Horizontal spatial discretization modularity within the spectral semi-implicit semi-Lagrangian ALADIN framework

ir. Steven Caluwaerts

Supervisors: prof. dr. P. Termonia, dr. ir. D. Degrauwe, prof. dr. H. Dejonghe
Dissertation submitted in fulfillment of the requirements for the degree of
Doctor (Ph.D.) in Science: Physics

Department of Physics and Astronomy
Faculty of Sciences
Ghent University
April 2016

UNIVERSITEIT
GENT

Contents

List of Figures

List of Tables

Used abbreviations

CFL	Courant-Friedrichs-Lewy
DFI	digital filter initialization
FD	finite-difference
FE	finite-element
HEVI	horizontally explicit vertically implicit
ICI	iteratively centered implicit
IGW	inertia-gravity waves
HPC	high performance computing
LAM	limited area model
LBC	lateral boundary condition
NWP	numerical weather prediction
RMS	root mean square
SI	semi-implicit
SL	semi-Lagrangian
SWE	shallow water equations
2TL	two timelevel

Acknowledgements

During the past years I was fortunate to get the opportunity to do research in atmospheric modeling, a topic that lies close to my heart. Today I am very proud to present some results in this PhD thesis. But first, I would like to thank some people whose support, help, or presence were necessary to attain this result.

I am very grateful to the support of my promoter team consisting of Piet, Daan, and Herwig. Piet was apart from my scientific guide in the NWP world also an excellent coach. I very much appreciate the chances and confidence he gave me, his vision, and ceaseless optimism. Despite his very busy agenda as ALADIN program manager, head of an RMI department, guest professor at Gent university,... I could always reach him if needed. He learned me a lot about atmospheric modeling but I also enjoyed his excellent management style. Daan's analytical and programming skills were useful to find the results I present in this thesis. I was fortunate to have such a patient, engaged, and pedagogical copromoter, and I also thank him for the good atmosphere in the RMI office. Finally, I would like to thank Herwig for his wise advice and the good cooperation we had during the past years.
My stays in Toulouse with the Météo-France colleagues Fabrice and Pierre often resulted in breakthroughs. I am grateful for their warm reception and constructive discussions, this was really very helpful for me. I also want to mention the role of Jean-Francois who unfortunately passed away last year. I never met a man with such a feeling for scientific problems, the value of discussions and mail exchanges with him often only became clear after several months...
The members of the thesis committee invested considerable time to read and comment my work. I would like to thank them because their remarks and the discussions with them substantially improved the manuscript.

I was lucky to be in a positive working environment and therefore I would like to express my thanks to my university and RMI colleagues. Special thanks for Daan, Flor, Geert, Gianfranco, Marjorie, and Waad who had to deal with me in their office. I am also grateful to my ping pong sparring partners Sebastien and again Flor.

A special thank you to my parents and family for their unconditional support during my studies and PhD. And last but very far from least I would like to thank my dearest Laure. She has always been there to support and, if needed, motivate me.

1 Introduction

Weather plays a very important role in our daily life. For a wide variety of sectors in the economy (e.g., agriculture, energy, insurance,...) accurate short and long term weather forecasts are of great economic value. For example, [9] discusses how to use wind forecasts to calculate the optimal flight routes in order to minimize the flight time and thus the fuel consumption and emissions. Today's modern society is highly vulnerable to extreme weather events that often result in huge human and economic losses. From 1970 to 2012 the total loss of human lives and the total economic damage due to weather disasters were estimated by [48] to be 1.94 million people and 2.4 trillion dollar respectively.

Apart from the capricious daily weather, the average state of our atmosphere - known as climate - and more specifically its evolution gained importance. As described in the *Impacts, Adaptation and Vulnerability* section of the fifth IPCC assessment report [38], climate change will have a large impact on the future of our planet and it is therefore considered as one of the major global challenges. Understanding the complex interactions that shape our climate and calculating projections of Earth's future climate based on emission scenarios became active research domains. The models used in these studies are largely based on existing weather models. As an example [46] recently studied climate projections over Europe by using ALADIN [2], a model with a long history in operational weather forecasting. Improvements and efficiency gains in weather modeling will therefore trigger further progress in climate research.

Due to its complex and multiscale nature, the only approach to predict the weather accurately consists of solving the equations that govern the motions in the atmosphere numerically. The history of weather modeling took off in the beginning of the 20th century. In 1904 Vilhelm Bjerkness realized that forecasting the weather is an initial value problem governed by the laws of physics [10]:

> If, as every scientifically inclined individual believes, atmospheric conditions develop according to natural laws from their precursors, it follows that the necessary and sufficient conditions for a rational solution of the problems of meteorological prediction are the following:

Parts of this chapter are published as [16].

1. The condition of the atmosphere must be known at a specific time with sufficient accuracy
2. The laws must be known, with sufficient accuracy, which determine the development of one weather condition from another.

This visionary idea still describes the main set-up of a current state-of-the-art atmospheric model where one starts from an analysis of meteorological observations to calculate future atmospheric states by integrating physical laws forward in time.
In 1922 Lewis Fry Richardson published a book titled *Weather prediction by numerical process* that outlined how to make an estimate of the future weather by integrating differential equations in time [85]. This brilliant work proposed to use a finite-difference method to solve a set of equations describing the motions in the atmosphere. It contained the results of one practical test case that unfortunately failed due to an initialization problem, as explained in detail in [63]. At the end of his book Richardson dedicated a chapter to some remaining problems illustrating his visionary mind. Firstly, he remarked that in order to make a good forecast one must start from an accurate initial state. At that time the observation network was very poor and clearly insufficient to make an accurate weather forecast. Richardson mentioned, for example, that information about water in clouds is often unknown. Even today the assimilation of meteorological data in a model is still a very active research domain. Secondly, he recognized that the complexity of the calculations made it unrealistic to expect a numerical weather forecast being finished before the forecast period was over. *'After so much hard reasoning, may one play with fantasy?'* he wrote, whereupon he gave the following description of a forecast factory:

> A myriad computers are at work upon the weather of the part of the map where each sits, but each computer attends only to one equation or part of an equation.

with *computers* at that time being humans doing computations. It is stunning to read in a book published in 1922 a description of a forecast factory corresponding closely to the current organization of weather forecasts on massively parallel supercomputers.
Less than 30 years later Richardson's dream came true when Charney, Fjörtoft, and Von Neumann performed a weather forecast successfully on the ENIAC, which stands for Electronic Numerical Integrator and Computer and was the first electronic general-purpose computer. The barotropic vorticity equations used for this experiment seriously simplified what happens in the atmosphere, but their work nevertheless proofed that by using an electronic computer one was able to make forecasts at the pace of the weather [18].
During the following decades computing power increased dramatically enabling ever more complex models to be evaluated at ever finer resolutions starting from ever more

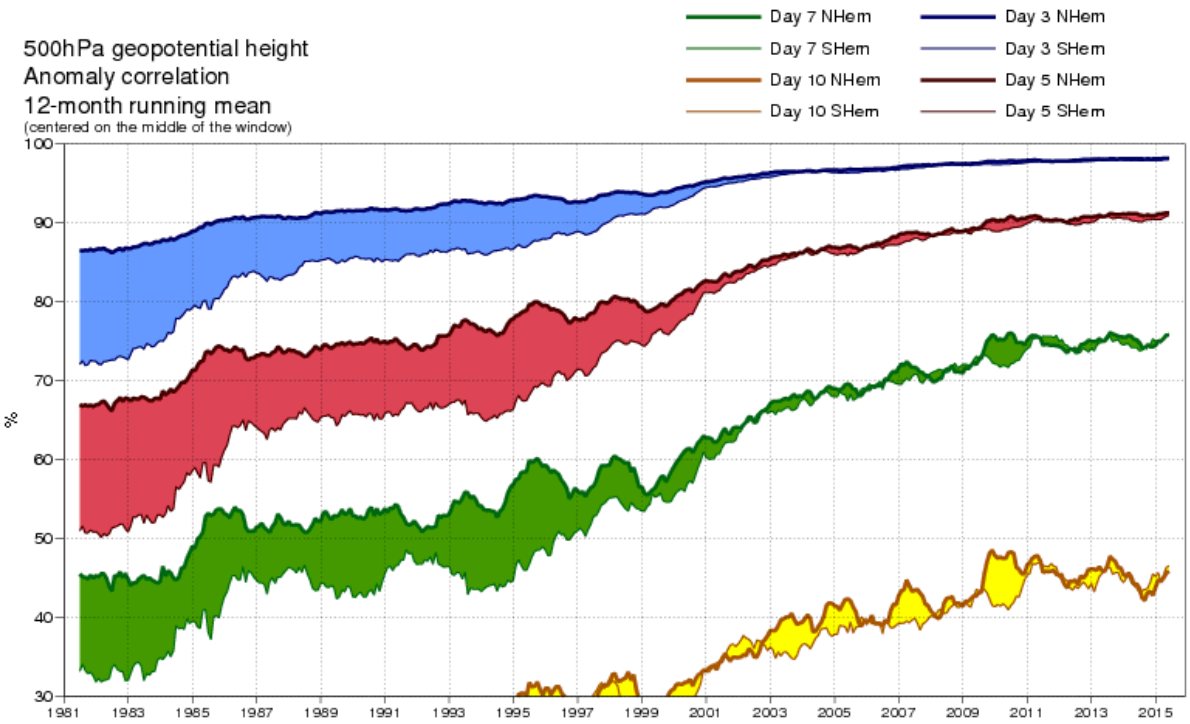


Figure 1.1: Time series of the annual running mean of anomaly correlations of HRES 500 hPa height forecasts evaluated against the operational analyses for the period January 1981 till present for different lead times (3, 5, 7, and 10 days). The score for each month is calculated over the 12-month period containing the 6 months before and the 5 months after this month. This figure is updated continuously on the ECMWF website (http://www.ecmwf.int/en/forecasts/charts/medium/anomaly-correlation-ecmwf-500hpa-height-forecasts).

accurate initial states. Figure 1.1, which shows the forecast scores of the high-resolution model run of the ECMWF (European Center for Medium range Weather Forecasts situated in Reading, UK) for the 500 hPa height[1], reveals the progress made last decades. The quality of the 3-day, 5-day, 7-day, and 10-day forecast is expressed by an anomaly correlation where 100% corresponds to a perfect forecast. The quality of the forecasts obviously improved a lot for all lead times. Based on Figure 1.1, one can estimate that the lead time at which the forecast exceeds a certain anomaly correlation increased by one day each decade. Improvements in the data assimilation, the model, and the observing systems were the main drivers of this quality increase [98].

Since accurate weather forecasts are imperative to society, the ambition of the numerical weather prediction (NWP) community must be to guarantee a continuous skill improvement for the next decades. This work can be considered as a tiny contribution to this huge challenge.

[1] This is the height above sea level where the pressure is equal to 500 hPa, typically a value around 5.5 km.

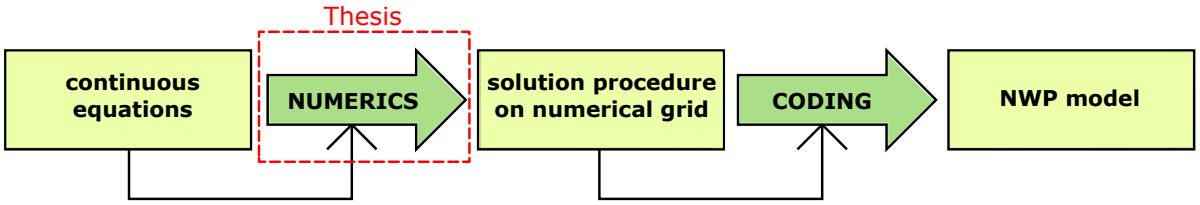


Figure 1.2: Schematic overview of the development of a NWP model.

1.1 How to develop a NWP model?

A NWP model consists of thousands of routines, which in total often exceed more than a million lines of computer code. This code could be seen as the end product of the model development. Figure 1.2 presents one point of view on NWP model development. It shows three different stages in the development of a NWP model:

1. First, one chooses a set of continuous equations describing the motions in the atmosphere. This choice will be related to the atmospheric processes one wants to resolve.
2. Second, one discretizes these equations by using numerical methods. This results in a numerical solution procedure. It is important that the discretization choices result in a stable and accurate scheme. The set of equations will influence the numerics used for the discretization.
3. Finally, one translates the discretized equations in computer code. The coding strategy will depend on the solution method that one constructed.

This thesis focuses on the numerics. The following sections introduce the equations used in NWP (section 1.2) and present an overview of the most used numerics (section 1.3). The coding aspects lie outside the scope of this thesis.

1.2 Continuous equations in NWP

The equations governing the motions in the atmosphere are derived by applying basic physics principles on an air parcel. [19] and [50] can be consulted for detailed derivations but here the focus is rather on the physical meaning of the equations. In vector notation the equations written out for a *coordinate system corotating with the Earth*, a

logical choice as we are living in such a system, are

$$\frac{d\mathbf{V}}{dt} = -2\boldsymbol{\Omega} \times \mathbf{V} - \frac{\nabla p}{\rho} - \mathbf{g} + \mathbf{Fr} \quad \text{(momentum equation)}, \tag{1.1}$$

$$\frac{dT}{dt} = \frac{RT}{C_p p}\frac{dp}{dt} + \frac{Q}{C_p} \quad \text{(thermodynamic energy equation)}, \tag{1.2}$$

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \mathbf{V} \quad \text{(continuity equation), and} \tag{1.3}$$

$$p = \rho R T \quad \text{(ideal gas law)}. \tag{1.4}$$

The variables and the physical meaning of the different terms in the previous equations are explained in the following paragraphs.
Equations (1.1), (1.2), and (1.3) are called *prognostic equations* because they describe the time evolution of the three-dimensional wind velocity vector $\mathbf{V}$, temperature T, and density ρ. These variables are called the prognostic variables. Equation (1.4) does not include any time derivation and is therefore called a *diagnostic equation*. The equations contain a *total time derivative* $\frac{d}{dt}$ that expresses the change in time while following the parcel. The total derivative does include the local rate of change $\frac{\partial}{\partial t}$ and *advection*

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \mathbf{V} \cdot \nabla f, \tag{1.5}$$

with f one of the meteorological variables. The advection term $\mathbf{V} \cdot \nabla f$ expresses the transport of the variable f by the wind field and plays an important role in NWP. Advection is one of the most important nonlinear terms in NWP [110] and, therefore, one of the main factors limiting the predictability of atmospheric motion.
The right hand side of the momentum equation (1.1) contains all forces, per unit mass, that are acting on an air parcel: the *Coriolis force* $-2\boldsymbol{\Omega} \times \mathbf{V}$ with $\boldsymbol{\Omega}$ the angular velocity vector of the Earth, the *pressure gradient force* $-\frac{\nabla p}{\rho}$ with pressure p, *apparent gravity* $-\mathbf{g}$ and friction $\mathbf{Fr}$. The Coriolis force is a consequence of the choice for a rotating coordinate frame and would not be present if the equations were written in an inertial reference frame. But there is a second fingerprint of the rotating coordinate frame. The *apparent* gravity $\mathbf{g}$ contains both the *true* Newtonian gravity of the Earth and the centrifugal force caused by the Earth's rotation, which depends on the latitude [80].
Different forms of the thermodynamic equation exist. Here the formulation (1.2) is chosen with C_p the specific heat at constant pressure and Q the diabatic heat transfer (e.g., latent heat released by phase changes of water). The continuity equation (1.3) expresses mass conservation. Finally, the ideal gas assumption (1.4), with $R = 287\,\mathrm{J\,K^{-1}\,kg^{-1}}$ being the ideal gas constant, relates pressure with density and temperature. The fluid flow equations can also be formulated in terms of other variables, for instance,

Exner pressure Π (a non-dimensionalized pressure) and *potential temperature* θ [2], which are used in the Met Office model described in [28].
In NWP models there will be some prognostic equations of the following form

$$\frac{dq}{dt} = S_q, \tag{1.6}$$

where q represents the mixing ratio[3] of a phase of water or a tracer (e.g., ozone) and S_q the source term for this component.
There are other important physical processes, such as radiation and microphysics, that should be included in order to develop a successful atmospheric model. Due to the small scales of these processes they are treated in a different way than the Euler equations. More information about the treatment of subgrid phenomena will follow in 2.2.1.
The set of equations (1.1) - (1.4) together with Eq. (1.6) describe all atmospheric motions and solving them properly is a *conditio sine qua non* to obtain a successful model. In what follows, we will assume a dry, adiabatic ($Q = 0$) atmosphere and study the impact of some commonly made approximations.

To perform an analytical study or to simplify the implementation of a discretization scheme one should introduce approximations in Eq. (1.1) - (1.4). Assumptions are typically based on a scale analysis of the magnitude of the different terms in an equation. However, one must realize that this will impact the solutions. If an approximation reduces a prognostic equation into a diagnostic one, it is called a filtering approximation. This is a far-stretching assumption because it turns out that it eliminates one wave solution. The previous could be useful, but one must realize that approximations limit the applicability of the model to the description of phenomena where these assumptions are actually valid. As a historical example, consider the first numerical weather forecast set up by Charney, Fjörtoft, and Von Neumann around 1950 [18]. The strongly simplified barotropic vorticity equation was used for this experiment and the motivation for this choice was twofold:

1. Given the limited availability of computing power at that time, they could only envision forecasting the barotropic evolution of the atmosphere.
2. As a result of the simplifications, the equations supported neither inertia-gravity nor sound waves releasing the strict time step limitations characteristic for such fast propagating modes. More information on this will follow in 1.3.1.

[2] The potential temperature θ of an air parcel is equal to the temperature the parcel would have if it is brought adiabatically to a certain reference pressure. This quantity is often used in meteorology.
[3] Mixing ratios are typically expressed as the mass of the component in question per mass of dry air.

Given the current computing facilities, today's NWP models are expected to predict, for instance, cyclogenesis and convection, accurately and therefore the barotropic vorticity equation is of no operational use anymore. Some of the most common approximations in NWP are discussed in the remaining part of this section.

Geopotential Φ is defined as $\nabla\Phi = -\mathbf{g}$, with $\mathbf{g}$ the earlier introduced apparent gravity. The surfaces of constant geopotential can be approximated as spheres. This is the so-called *spherical geopotential approximation* as explained in, for example, [132]. This is a popular approximation because it reduces the complexity of the equations when written in a spherical coordinate system. However, it is a systematic approximation and some wonder whether the cumulative effects are still negligible in, for example, climate modeling [108]. More accurate alternatives, like the *ellipsoidal geopotential approximation*, were therefore proposed and their impact was tested for simplified equation sets [8]. Today, these alternatives have not yet been applied operationally in NWP models.
As detailed in [50], by adopting a spherical coordinate system (λ,ϕ,r) with longitude λ, latitude ϕ, and r the distance from the center of the Earth, illustrated in Figure 1.3, the momentum equation (1.1) can be expanded in its components

$$\frac{du}{dt} = \frac{uv\tan\phi}{r} - \frac{uw}{r} - \frac{1}{\rho r\cos\phi}\frac{\partial p}{\partial\lambda} + 2\Omega v\sin\phi - 2\Omega w\cos\phi, \tag{1.7}$$

$$\frac{dv}{dt} = \frac{u^2\tan\phi}{r} - \frac{vw}{r} - \frac{1}{\rho r}\frac{\partial p}{\partial\phi} - 2\Omega u\sin\phi, \text{ and} \tag{1.8}$$

$$\frac{dw}{dt} = \frac{u^2+v^2}{r} - g - \frac{1}{\rho}\frac{\partial p}{\partial r} + 2\Omega u\cos\phi. \tag{1.9}$$

The wind vector $\mathbf{V} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ consists of three components, with $\mathbf{i}$ the eastward pointing unit vector along a latitude circle, $\mathbf{j}$ the northward pointing unit vector along a meridian, and $\mathbf{k}$ the unit vector pointing upwards along the local vertical. The unit vectors' dependency on the position on the Earth explains the so-called curvature terms consisting of products of the velocity components. Equations (1.7) - (1.9) together with the continuity, thermodynamic energy, and ideal gas equation are called the *deep atmosphere Euler equations*.

A further approximation assumes a *shallow atmosphere* by replacing the radial distance r by the radius a of the Earth. For reasons of conservation, this shallowness comes together with two subsidiary approximations:

- neglecting some of the curvature terms and

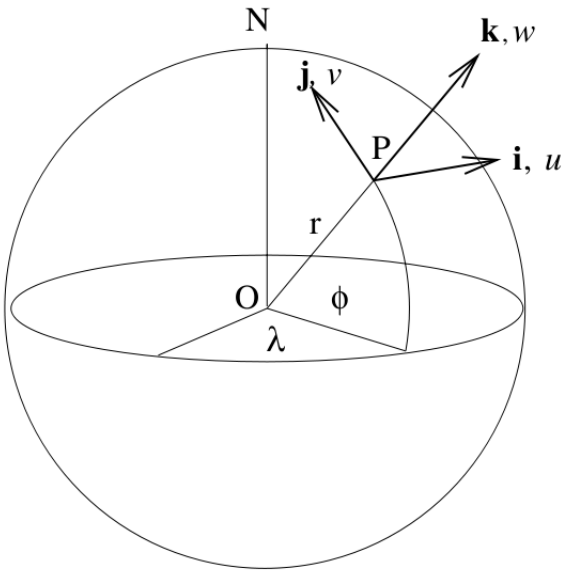


Figure 1.3: Spherical coordinate system as it is used often in NWP, with O the center of the Earth and N the North Pole. At each position, specified by latitude ϕ, longitude λ, and radius r, an orthogonal basis of unit vectors $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ can be defined. This figure is copied from [132].

- neglecting the Coriolis terms containing $\cos\phi$,

all together known as the *traditional approximation* [36]. Introducing these simplifications into Eq. (1.7), Eq. (1.8), and Eq. (1.9) yields

$$\frac{du}{dt} = \frac{uv\tan\phi}{a} - \frac{1}{\rho a\cos\phi}\frac{\partial p}{\partial\lambda} + 2\Omega v\sin\phi, \tag{1.10}$$

$$\frac{dv}{dt} = \frac{u^2\tan\phi}{a} - \frac{1}{\rho a}\frac{\partial p}{\partial\phi} - 2\Omega u\sin\phi, \text{ and} \tag{1.11}$$

$$\frac{dw}{dt} = -g - \frac{1}{\rho}\frac{\partial p}{\partial r}. \tag{1.12}$$

The previous scalar momentum equations together with the continuity, thermodynamic energy, and ideal gas equation are denoted as the *shallow atmosphere Euler equations*. The shallow atmosphere approximation is omnipresent in NWP although keeping the deep atmosphere has some advantages, as explained in [132]:

> Finally, we would reiterate that experience with deep models suggests that they give very similar results to shallow models under the currently prevailing terrestrial conditions. The value of deep models lies in their comprehensiveness. They apply to all planetary atmospheres for which the spherical geopotential approximation is valid, as well as to the terrestrial atmosphere even under conditions of much weaker static stability than

currently prevailing.

Today, the Met Office is using the deep atmosphere equations operationally [28] while other communities, for instance ARPEGE, investigate this option [139].

Last decades, nearly all global NWP models were based on the *hydrostatic primitive equations*. This equation set is obtained by applying a filtering approximation to Eq. (1.12). Assuming negligible accelerations along the vertical ($\frac{dw}{dt} \approx 0$) gives rise to the hydrostatic equation

$$\frac{\partial p}{\partial r} = -\rho g. \tag{1.13}$$

The previous equation expresses that the pressure difference between two vertical levels is equal to the weight of the layer of air in between[4]. Figure 1.4 illustrates the impact of the hydrostatic assumption on the wave solutions present in a statically stable isothermal atmosphere [4]. The figure presents the dispersion relation, that is the frequency ω of an atmospheric wave in function of its wavenumber k for different vertical wavenumbers m. A more detailed discussion about dispersion relations follows in 3.1.2 but some key elements are already present in Figure 1.4:

- The *acoustic* or *sound* waves are eliminated by the hydrostatic assumption except for the horizontally propagating ($m = 0$) sound wave, called the Lamb wave.[5] As can be seen in the upper plot of Figure 1.4, the acoustic waves have the largest frequency and thus phase velocity $\frac{\omega}{k}$. By filtering the vertically propagating acoustic waves, which do contain little energy and do not have any meteorological relevance, the stability constraints become less stringent as will be explained later. The presence of the Lamb wave in the hydrostatic problem does not pose a numerical problem since its dispersion relation is very close to the one of the fastest inertia-gravity waves.
- The dispersion relation of the *inertia-gravity waves* (IGWs) under the hydrostatic assumption is given by

$$\omega_{IG}^2 = f^2 + \frac{N^2}{m^2 + 1/\left(4H_0^2\right)}k^2, \tag{1.14}$$

  as written in [4], with N the Brunt-Väissälä frequency (related to the static stability of the atmosphere) and $H_0 = \frac{RT_0}{g}$ the scale height of the atmosphere. In the long wave limit, the frequency is identical to the nonhydrostatic one. But, as

[4] For the sake of completeness, one should remark that the hydrostatic approximation can also be combined with the deep atmosphere equations resulting in the so-called *quasi-hydrostatic equations* [132].

[5] The hydrostatic approximation does only change the vertical momentum equation, one could intuitively understand that a horizontal sound wave is not affected by this approximation.

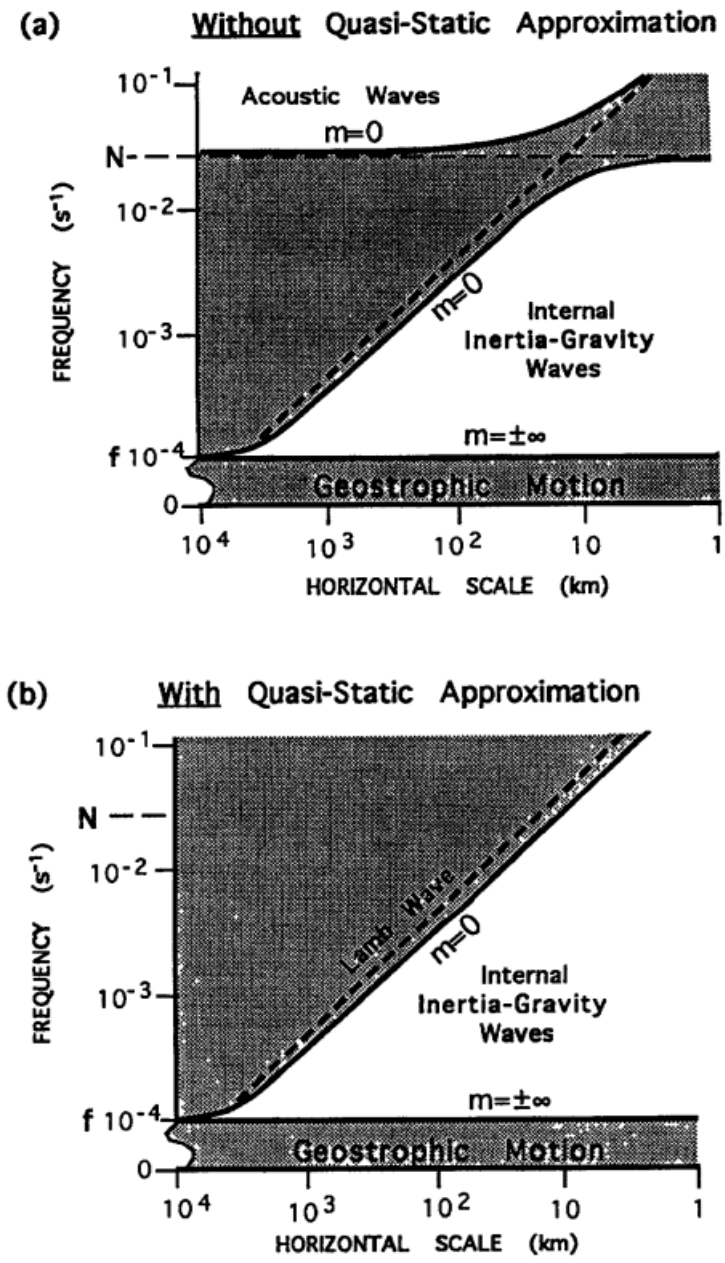


Figure 1.4: Impact of the hydrostatic approximation (here called quasi-static approximation) on the wave modes of the model. The frequencies correspond to small-amplitude perturbations in an isothermal ($T_0 = 250\,\mathrm{K}$), statically stable atmosphere in rest, with Coriolis constant $f = 10^{-4}\,\mathrm{s}^{-1}$ and vertical wavenumber m. Figure is taken from [4].

can be seen in Figure 1.4, for the short wave limit a behavior different from the nonhydrostatic one, given by

$$\omega_{IG,SW}^2 \approx \frac{N^2}{m^2 + k^2 + 1/\left(4H_0^2\right)} k^2, \tag{1.15}$$

is found [74]. This shows that the applicability of the hydrostatic assumption is limited to certain scales.

- The *Rossby mode* is not changed by the hydrostatic approximation. This is important as Rossby waves are the main driver for synoptic scale meteorology[6]. Figure 1.4 is based on an analysis with a constant Coriolis parameter f, which results in a stationary Rossby mode corresponding to geostrophic balance.

To conclude, the hydrostatic approximation is an elegant way to eliminate vertically propagating sound waves, which are otherwise imposing a strict time step limitation despite their meteorological irrelevance. Given that it only distort IGWs with short horizontal scales, the hydrostatic primitive equations are a popular equation set for synoptic scale models.

These days, there is a trend towards *seamless models*, also known as *unified models*, in the atmospheric modeling communities [14]. These models are developed to predict atmospheric evolution over all spatial and temporal scales, hence they could serve high-resolution mesoscale modeling, synoptic scale global modeling, and even long term climate modeling. This evolution towards seamless models is mainly driven by the huge investment and maintenance costs of models that make it far more efficient to have only one atmospheric model code for all scales and applications.
The use of a set of equations that describes atmospheric phenomena over all scales is crucial for a successful seamless model. The hydrostatic models are no candidate because the assumption $\frac{dw}{dt} \approx 0$ is no longer valid for phenomena with small scales (roughly below 10 km). In literature two alternatives are proposed:

1. A scientifically straightforward option would be to solve the *Euler* (also called *fully elastic* or *compressible*) equations without any approximation. However, this means that the numerics need to deal with fast propagating sound waves.
2. Many sets of equations (e.g., anelastic equations,...) were proposed that retain the nonhydrostatism in order to admit adequate mesoscale modeling but use a different filtering approximation to eradicate sound waves. Because acoustic

[6] Synoptic scale meteorology treats weather phenomena with horizontal lengths ranging from about 1000 km to 5000 km e.g., mid-latitude cyclones. A graphical overview of scales of atmospheric motions is presented in Figure 2.2

waves are characterized by oscillating density fields, most alternatives are based on approximations in the continuity equation (1.3) by, for example, dictating incompressibility $\nabla \cdot \mathbf{V} = 0$. However, with a linearized normal mode analysis it was shown in [27] that these approaches distort the Rossby wave propagation making them inappropriate for global scale modeling[7]. On the other hand, nonlinear numerical tests in [102] suggest that these differences are mainly important for NWP and may have less impact on long term climate runs. Recently, progress was reported in the search for equation sets filtering the sound waves without deteriorating the Rossby waves [5] [33].

To conclude, one can synthesize atmospheric modeling symbolically by the following initial value problem

$$\frac{\partial \mathbf{U}(t)}{\partial t} = \mathbf{A}(\mathbf{U}(t)) + \mathbf{M}(\mathbf{U}(t)) + \mathbf{P}(\mathbf{U}(t)) \text{ and } \mathbf{U}(t=0) = \mathbf{U_0}, \tag{1.16}$$

with $\mathbf{A}(\mathbf{U}(t)) = -\mathbf{V} \cdot \nabla \mathbf{U}(t)$ representing advection, $\mathbf{M}$ representing the other resolved forcings (pressure gradient force, Coriolis force and gravity), and $\mathbf{P}$ the tendency of the *physics parameterizations* of unresolved processes, as explained in more detail in 2.2.1. Advection is written out separately in Eq. (1.16) because it is the dominant nonlinear process in weather modeling and therefore needs special attention. The used approximations and the resolution of the model will determine the exact form of the operators $\mathbf{M}$ and $\mathbf{P}$. Remark that Eq. (1.16) should be solved by applying appropriate vertical (and, in case of a limited area model, lateral) boundary conditions on the domain.

The physics tendencies $\mathbf{P}(\mathbf{U}(t))$ are assumed to be known explicitly in Eq. (1.16). Numerical integration of the so-called *dynamical core*,

$$\frac{\partial \mathbf{U}(t)}{\partial t} = \mathbf{A}(\mathbf{U}(t)) + \mathbf{M}(\mathbf{U}(t)), \tag{1.17}$$

in an efficient and stable way while approaching as closely as possible the analytical properties of the different atmospheric waves forms the main challenge of the *numerics*.

1.3 Numerics in NWP

Following the presentation of the continuous partial differential equations and some of its most common approximations in the previous section, we now provide an in-

[7] These equations could be used for mesoscale modeling, because the larger scale Rossby waves are mainly determined by the coupling (hydrostatic) model.

troduction to the discretization of Eq. (1.17). The discretization procedure, which will result in a numerical scheme, forms a crucial step in the development of a successful numerical weather model.
The numerical scheme will consist of both a time discretization and space discretization as **A** and **M** contain spatial derivatives. The spatial operators act along the horizontal and vertical direction and one thus ends up with a 3D problem. If the vertical operations do not depend on the horizontal position, this 3D numerical problem can be split into a complementary but rather independent 2D horizontal and 1D vertical problem. The scalability problems of the dynamical core mainly occur in the horizontal dimension, we thus focus on the horizontal spatial discretization and touch the vertical discretization only briefly[8]. The presented methods ought to give a first introduction to this broad topic. The reader is referred to textbooks (e.g [11, 19, 34]) for more exhaustive overviews of discretization approaches in NWP.
Table 1.1 presents the discretization choices made in some of the European NWP models that are used operationally with ongoing research and development. The table illustrates that a broad spectrum of methods is used. All methods mentioned in Table 1.1 will be explained later on in this section.

1.3.1 Time discretization

Two categories of time discretization schemes are used in NWP: explicit and semi-implicit schemes. In a scheme based on an *explicit* time discretization all forcing terms (right-hand side of Eq. (1.17)) are evaluated at previous time levels, resulting in a direct update problem. By evaluating the forcing terms both at previous times and at the current instant of time, one obtains an *implicit* scheme, which results in an elliptic problem. The choice of the time discretization scheme has an impact on:

1. the propagation properties of wave modes, for instance waves are slowed down by a semi-implicit treatment.
2. the numerical stability of the scheme, for instance it is well known that an explicit time discretization results in a necessary condition to obtain a stable scheme. This so-called Courant-Friedrichs-Lewy (CFL) criterion [24] reads $V\frac{\Delta t}{\Delta x} < C$ with V the velocity, Δx the shortest grid distance on the mesh, Δt the time step, and C a constant depending on the discretization scheme used. The fastest propagating mode will determine the time step restriction of the dynamical core. Finer model resolutions will make this time step criterion more strict.

[8] In atmospheric models that are based on an explicit time discretization along the vertical dimension, the high density of vertical levels close to the surface may pose a strict CFL criterion and thus another scalability problem. We come back on this when we explain the HEVI (horizontally explicit vertically implicit) approach.

Table 1.1: Overview of the discretization choices made in some European NWP models that are operationally in use. The used abbreviations: 2 TL = two time level, SISL = semi-implicit semi-Lagrangian, FD = finite-difference and FV = finite-volume.

	Time discretization	Horizontal space discretization / grid
ICON [140] (DWD)	2 TL predictor-corrector HEVI	FD-FV on icosahedral-triangular Arakawa C-grid
IFS /ARPEGE/ALADIN [2, 116, 131] (ECMWF/Météo-France/ALADIN)	2 TL SISL	spectral on reduced Gaussian grid, stretched grid, or rectangular grid (collocated variables)
ENDGame [28, 40] (Met Office)	2 TL SISL (predictor-corrector for NL terms)	FD on lat-lon Arakawa C-grid
SL-AV [124, 125] (Russia Hydrometeor)	2 TL SISL	partly FD, partly spectral on lat-lon grid (collocated variables)

3. the order of accuracy, for instance an evaluation at more time levels may result in a higher order of accuracy. However, this will also necessitate more data storage. In NWP the time step is mostly determined by stability and not by accuracy, which explains why often two timelevel (2TL) methods with only first or second-order accuracy are used.

The straightforward explicit time discretization of Eq. (1.17) would be the *forward* scheme (also called Euler's method) written as

$$\mathbf{U}^{+} = \mathbf{U}^{0} + \Delta t \left[\mathcal{A}\left(\mathbf{U}^{0}\right) + \mathcal{M}\left(\mathbf{U}^{0}\right) \right], \tag{1.18}$$

with $\mathbf{U}^0$ and $\mathbf{U}^+$ the vectors containing the prognostic variables in all grid points at time t^0 and $t^+ = t^0 + \Delta t$, respectively. The spatially discretized operators corresponding to $\mathbf{A}$ and $\mathbf{M}$ are represented by $\mathcal{A}$ and $\mathcal{M}$. However, this scheme is in general unconditionally unstable, as derived for example in [19], meaning that the amplitude of physically nonamplifying waves will be amplified numerically by this scheme, independent of the choice of the time step. However, a forward scheme combined with an upwind spatial discretization of advection is stable and forms the basis of the MP-DATA approach e.g., [102].
In NWP other explicit temporal discretization methods are used, for example the two-

timelevel *predictor-corrector*, also known as the second-order *Runge-Kutta* scheme or the *Heun* scheme. This scheme reads as

$$\mathbf{U}^* = \mathbf{U}^0 + \Delta t \left[\mathcal{M}\left(\mathbf{U}^0\right) + \mathcal{A}\left(\mathbf{U}^0\right) \right] \text{ and} \tag{1.19}$$

$$\mathbf{U}^+ = \mathbf{U}^0 + \frac{\Delta t}{2} \left[\mathcal{M}\left(\mathbf{U}^0\right) + \mathcal{M}\left(\mathbf{U}^*\right) + \mathcal{A}\left(\mathbf{U}^0\right) + \mathcal{A}\left(\mathbf{U}^*\right) \right]. \tag{1.20}$$

The predictor step (1.19) uses a forward Euler scheme to find a preliminary updated state vector $\mathbf{U}^*$. During the corrector step, Eq. (1.20), $\mathbf{U}^+$ is calculated by interpolating the forcing over $\mathbf{U}^0$ and $\mathbf{U}^*$. This scheme could be evaluated easily in the sense that one does not need to solve elliptic problems. This approach is still unconditionally amplifying [34], despite having an amplification that is much weaker than the one of the forward Euler scheme. The ICON model [140], which is based on a dynamical core solving the Euler equations with a 2TL scheme, uses a stable variant of such a scheme. As explained earlier, the time step of an explicit scheme is restricted by the fastest propagating wave. However, for the stability of other waves moving at lower speeds these short time step integrations are overly strict and therefore the *split-explicit* scheme was developed. In this approach different terms in the equations are integrated with different time steps. The terms responsible for the fastest propagating waves are integrated with a smaller time step than that part of the equations governing the slower waves. In this way one decreases the number of evaluations of the *slow dynamics* terms. This approach was first used for the time discretization in a model developed to study storm dynamics where it was used to integrate the terms responsible for the fast acoustic waves with a smaller time step than the remaining terms [57]. A similar approach is still today used in the WRF dynamical core, documented in [101].
Two remarks should be made regarding split-explicit methods:

1. Splitting the terms into fast and slow waves is often not possible because the same term could be involved both in slow and fast wave propagation. Therefore, a careful analysis, as performed for example in [44], is needed when a split-explicit scheme is used.

2. The ICON model, which solves the nonhydrostatic Euler equations, does not use a smaller time step for the acoustic waves. This decision was based on the observation that high up in the stratosphere IGWs were found with speeds close to the sound speed [140].

As an alternative to explicit time integration schemes, one can average the forcing terms over the timelevels t^0 and t^+. This slows down the fastest waves and permits the use of longer time steps. These schemes, which were introduced in NWP by Robert [92], are called *implicit* time discretization schemes and result in an elliptic

problem:

$$\mathbf{U}^{+} = \mathbf{U}^{0} + \frac{\Delta t}{2}\left[\mathcal{A}\left(\mathbf{U}^{0}\right) + \mathcal{A}\left(\mathbf{U}^{+}\right) + \mathcal{M}\left(\mathbf{U}^{0}\right) + \mathcal{M}\left(\mathbf{U}^{+}\right)\right]. \tag{1.21}$$

Equation (1.21) is not used as such in NWP because of its complexity. The operators $\mathcal{A}$ and $\mathcal{M}$ contain, indeed, nonlinear terms making an implicit treatment very cumbersome, except if one would turn to iterative methods. Therefore, the dynamics operator is split into a linear operator $\mathcal{L}$ and a nonlinear residual $\mathcal{M} - \mathcal{L}$ and only the linear part is treated in an implicit way. This yields

$$\mathbf{U}^{+} = \mathbf{U}^{0} + \frac{\Delta t}{2}\left[\mathcal{L}\left(\mathbf{U}^{0}\right) + \mathcal{L}\left(\mathbf{U}^{+}\right) + 2\mathcal{A}\left(\tilde{\mathbf{U}}\right) + 2\left(\mathcal{M} - \mathcal{L}\right)\left(\tilde{\mathbf{U}}\right)\right]. \tag{1.22}$$

The previous approach, which treats part of the forcing in an implicit and part in an explicit way, is called the *semi-implicit* (SI) or *implicit-explicit* (IMEX) method [35] and results in an Helmholtz problem

$$\mathcal{H} f^{+} = R, \tag{1.23}$$

with f^{+} one of the prognostic fields at t^{+} and $\mathcal{H}$ the 3D Helmholtz operator. Depending on how the linearized operator $\mathcal{L}$ is constructed, three different types of SI Helmholtz problems are distinguished [7]:

1. the linearization is done around a constant in time, horizontally homogeneous reference state and one obtains $\mathcal{H}_0$,
2. the reference state is constant but horizontally inhomogeneous resulting in $\mathcal{H}\left(x, y\right)$, and
3. a time dependent, horizontally inhomogeneous reference state is assumed resulting in $\mathcal{H}\left(x, y, t\right)$.

From the numerical point of view the so-called constant coefficient Helmholtz operator (option 1) results in the easiest Eq. (1.23) to solve. However, the reference state used for the linearization could be quite different from the real state of the atmosphere. This will be translated into large nonlinear residuals $\left(\mathcal{M} - \mathcal{L}\right)\left(\tilde{\mathbf{U}}\right)$ and, because these are treated explicitly, that could make the scheme unstable. This is illustrated in [96] that investigated the impact of the thermal nonlinear terms, which are a consequence of the differences between the reference and real temperature profile. Apart from the thermal terms, the constant coefficients approach necessitates an explicit treatment of the orography terms, which are also vulnerable for instabilities [55]. While developing a new nonhydrostatical ALADIN kernel, Bénard studied in a series of papers the stability of such constant coefficient SI schemes. He concluded that, by a good choice for the prognostic variables and reference state, constant coefficients schemes are suitable for

mesoscale NWP [6].
The Helmholtz operators corresponding to the constant (option 2) or time dependent (option 3), horizontally inhomogeneous reference state result in more complex, truly 3D problems. A well known illustration of the latter is the nonhydrostatic model presented by Skamarock in [99]. They linearize around the state of the atmosphere at the previous time step thereby minimizing the nonlinear residuals at the expense of obtaining a time varying, inhomogeneous reference state. This is claimed to provide greater accuracy and stability, especially over steep orography. This methodology is used, for example, in the ENDGame model developed by the Met Office[9] [28, 40].

Analysis shows that the waves described by the implicitly treated part $\mathcal{L}$ are significantly slowed down, so that the linear part $\mathcal{L}$ is unconditionally stable. By choosing $\mathcal{L}$ such that it contains the terms responsible for the fastest moving waves, for example, the sound waves in fully compressible equations, one can weaken the CFL-time step criterion. The method was first illustrated in NWP in 1971 by Kwizak and Robert [58] where it was used to slow down the short scale gravity waves. This allowed the time step to increase by a factor 6. The nonlinear part of the dynamics operator and the physics can be evaluated in multiple ways, symbolized by $\tilde{\mathbf{U}}$ in Eq. (1.22). A predictor-corrector approach is used to incorporate the nonlinear terms in the ENDGame model [28] and the nonhydrostatic version of ALADIN [7]. Hortal on the other hand proposed, within the context of the IFS model, to use a stable extrapolation of the nonlinear terms [52]. Remind that the stability of the implicitly treated part is no guarantee that the scheme as a whole will be stable, as explained in [96].
The *horizontally explicit vertically implicit* (HEVI) scheme can be considered as part of the SI family of schemes. As explained in 1.2, there is a trend in NWP towards unified modeling meaning that one model is used to simulate atmospheric phenomena over all scales [14]. For this reason the fully compressible equations are gaining importance [27]. But this set of equations does permit 3D sound waves, which do not have any meteorological relevance, to propagate at high speeds. In NWP models the vertical grid spacing is strongly dependent on the height. High-up in the atmosphere the resolution is low, but close to the surface the vertical grid spacing can be of the order of meters, much smaller than the horizontal resolution. If treated explicitly, vertically propagating sound waves would necessitate a very short time step. Therefore, the HEVI approach, which is an explicit scheme treating only the terms for vertical sound

[9] Until 2014 a version called *New Dynamics*, described in detail by Davies, was used [28]. However, the model needed quite some damping to control instabilities and the 3D Helmholtz solver turned out to be insufficiently scalable and therefore an updated version, *ENDGame*, is currently used. ENDGame is still based on the same equation set and numerics, but the equations are solved in a more iterative approach. The changes are described in a Met Office documentation report [40].

wave propagation implicitly, is gaining importance. The ICON model, described in [140], is a recent example of a HEVI scheme.

Robert [93] estimated that even in SI schemes the spatial discretization errors are still 40 times larger than the temporal discretization errors. Thereby the time step is still limited by stability constraints and not by accuracy. Advection processes were found to limit the maximum stable time step. Apart from the direct evaluation of the nonlinear advection terms, which was used these days in NWP, a Lagrangian alternative, where advection was evaluated by following parcels along their trajectory, was known [41]. Because this Lagrangian approach is difficult to combine with a fixed grid, Sawyer proposed a semi-Lagrangian (SL) approach where trajectories were calculated for parcels on a uniform grid [95]. Robert showed that combining the SL treatment for advection with a SI discretization permits stable integrations of the equations with even longer time steps [93]. SISL schemes can be represented symbolically as

$$\mathbf{U}^{+} = \mathbf{U}_D^0 + \frac{\Delta t}{2}\left[\mathcal{L}\left(\mathbf{U}_D^0\right) + \mathcal{L}\left(\mathbf{U}^{+}\right) + 2\left(\mathcal{M} - \mathcal{L}\right)\left(\tilde{\mathbf{U}}\right)\right], \tag{1.24}$$

wherein there is no longer an evaluation of the nonlinear advection term $\mathcal{A}$. Evaluations at time t^0 are undertaken in the departure points, which are found by tracking back the positions at time t^0 of parcels that are at time t^+ in the grid points of the fixed grid. An evaluation in a departure point is denoted by the subscript $_D$. The nonlinear terms are evaluated along the SL trajectory but different methods exist. The SL efficiency gain by the reduction of the number of time integrations is partly offset by the trajectory calculations and the extra communications needed to do the interpolations to departure points. Advection terms are the most important nonlinear terms [110]. Treating them in a SL way removes the main source for aliasing and thereby the need to filter the shortest waves of the model [52].
Due to their enhanced stability, SISL schemes became popular time discretization schemes for atmospheric modeling. Three out of the four models presented in Table 1.1 are based on the SISL method.

Before closing this introduction about temporal discretization schemes, it is interesting to note the recent attention for the *modified equation* approach [131, 102, 103]. Characteristic for such a scheme is the increase in the order of accuracy by subtracting part of the truncation error from the original equation. Solving the modified equation will then lead to a higher order of accuracy. The earlier cited literature illustrates that the modified equation approach can be used to construct 2TL scheme with second order accuracy for both hydrostatic and non-hydrostatic equations. It will be explained that

modularity is an important constraint for this thesis. As this method would lead to, for example, a timestep organization quite different from the current one, we will not study in detail the modified equation approach for this thesis.

1.3.2 Horizontal space discretization and grid choice

This subsection presents an overview of horizontal spatial discretization strategies used in this thesis. For more information the reader is referred to the earlier mentioned textbooks of Durran [34] and Coiffier [19]. Remark that different discretization strategies can lead to the same discretized equation.

The finite-difference method

The most intuitive way for the numerical representation of spatial operators is the *finite-difference* (FD) approach. The FD method approximates the derivative of a function in a certain point by a weighted subtraction of the values of the function in neighboring points. This methodology is justified by the definition of the derivative of a differentiable function f in a point x_i:

$$\frac{\partial f}{\partial x}(x_i) = \lim_{\epsilon \to 0} \frac{f(x_i + \epsilon) - f(x_i)}{\epsilon}. \tag{1.25}$$

The one-dimensional constant advection equation reads:

$$\frac{\partial u}{\partial t}(x,t) = -C\frac{\partial u(x,t)}{\partial x}, \quad \forall x \in \Omega \tag{1.26}$$

with u being a function advected with a constant velocity C and Ω the domain where the problem is defined. Equation (1.26) is often used to illustrate discretization schemes. Hereafter, different FD approximations for the tendency of the function u in the i^{th} grid point are given

$$\left(\frac{\partial u}{\partial t}\right)_i = -C\frac{u_i - u_{i-1}}{\Delta x} \quad (1^{\text{st}}\text{-order accurate FD}), \tag{1.27}$$

$$\left(\frac{\partial u}{\partial t}\right)_i = -C\frac{u_{i+1} - u_i}{\Delta x} \quad (1^{\text{st}}\text{-order accurate FD}), \tag{1.28}$$

$$\left(\frac{\partial u}{\partial t}\right)_i = -C\frac{u_{i+1} - u_{i-1}}{2\Delta x} \quad (2^{\text{nd}}\text{-order accurate FD}), \text{ and} \tag{1.29}$$

$$\left(\frac{\partial u}{\partial t}\right)_i = -C\left[\frac{4}{3}\frac{u_{i+1} - u_{i-1}}{2\Delta x} - \frac{1}{3}\frac{u_{i+2} - u_{i-2}}{4\Delta x}\right] \quad (4^{\text{th}}\text{-order accurate FD}). \tag{1.30}$$

For reasons of simplicity, a constant grid distance Δx was assumed although the previous expressions can be modified for a variable resolution mesh. The order of accuracy of the approximations (1.27) - (1.30) is derived by replacing the values of u by a the Taylor expansion around u_i

$$u_{i+j} = u(x_i + j\Delta x) = u_i + (j\Delta x)\left(\frac{\partial u}{\partial x}\right)_i + \frac{(j\Delta x)^2}{2}\left(\frac{\partial^2 u}{\partial x^2}\right)_i + ... \tag{1.31}$$

The order of accuracy of the discretization can be increased by including more points in the FD approximation. Similar expressions can be derived for second-order derivatives, third-order derivatives,...
FD schemes are often combined with a *staggered grid* formulation where different variables are evaluated in different grid points [68]. The motivation for staggered grids will be clarified in chapter 3.
The NWP scheme proposed by Richardson [85] was based on a second-order accurate FD spatial discretization. Ever since, schemes based on FD approximations for horizontal derivatives are commonplace in NWP. Today, the SL-AV model [124] and the ENDGame model [14] are illustrations of FD spatial discretizations in the European NWP community.

Galerkin methods

A second category of spatial discretization schemes, called Galerkin methods, is based on an *expansion* in terms of *basis functions*. Let us demonstrate this approach by again solving numerically the 1D constant advection equation. The method then assumes that the numerical solution $u_N(x,t)$ of Eq (1.26) can be written as an expansion over N basis functions $\phi_i(x)$ defined in Ω:

$$u_N(x,t) = \sum_{i=1}^{N} U_i(t)\phi_i(x), \tag{1.32}$$

where $U_i(t)$ represents the weight of the basis function ϕ_i. The modeling problem is now translated into finding the evolution in time of the N expansion coefficients $U_i(t)$. Substitution of Eq. (1.32) into the advection equation (1.26) yields:

$$\sum_{i=1}^{N} \frac{dU_i}{dt}(t)\,\phi_i(x) = -C\sum_{i=1}^{N} U_i(t)\frac{d\phi_i}{dx}(x). \tag{1.33}$$

Using Eq. (1.33), one could define the residual R as:

$$R(x,t) = \sum_{i=1}^{N} \frac{dU_i}{dt}(t)\,\phi_i(x) + C\sum_{i=1}^{N} U_i(t)\frac{d\phi_i}{dx}(x). \tag{1.34}$$

One must strive towards a minimal R. If $R(x,t) = 0$ for each position x at each time t, the proposed solution $u_N(x,t)$ is an exact solution of Eq. (1.26).
Different strategies for the minimization of the residual will result in different coefficients U_i. One could, for example, use the *collocation* method that dictates the residual to be exactly equal to 0 in N grid points. However, the preferred method in NWP is the *Galerkin* method, which is based on the evaluation of the residual against the N basis functions $\phi_j(x)$ over the whole domain Ω:

$$\int_\Omega \phi_j(x)\,R(x,t)\,dx = 0.\ \ \forall j = 1,2,...,N \tag{1.35}$$

By using Eq. (1.34), one gets:

$$\sum_{i=1}^{N} M_{ji}\frac{dU_i}{dt}(t) = -C\sum_{i=1}^{N} D_{ji}U_i(t),\ \ \forall j = 1,2,...,N \tag{1.36}$$

with

$$M_{ji} = \int_\Omega \phi_i(x)\,\phi_j(x)\,dx \text{ and} \tag{1.37}$$

$$D_{ji} = \int_\Omega \frac{d\phi_i}{dx}(x)\,\phi_j(x)\,dx. \tag{1.38}$$

Equation (1.36) is often summarized by the following matrix equation

$$\mathbf{M}\times\frac{d\mathbf{U}}{dt} = \mathbf{D}\times\mathbf{U}, \tag{1.39}$$

with $\mathbf{M}$ the $N\times N$ mass matrix, $\mathbf{D}$ the $N\times N$ differentiation matrix, and $\mathbf{U}$ the $N\times 1$ column vector containing the coefficients of the basis functions. Whether these matrices are sparse or full matrices will depend on the chosen set of basis functions $\phi_i(x)$. By choosing for example a set of orthogonal basis functions $\phi_i(x)$, the mass matrix $\mathbf{M}$, consisting of the coefficients M_{ji}, becomes diagonal:

$$M_{ji} = \int_\Omega \phi_i(\mathbf{r})\,\phi_j(\mathbf{r})\,d\mathbf{r} = \delta_{ij},\ \ \forall i,j = 1,2,...,N, \tag{1.40}$$

and Eq. (1.36) decouples into a set of uncoupled (or direct) update equations for the coefficients U_i. In general, the mass matrix $\mathbf{M}$ contains off-diagonal elements and the

prognostic equations (1.36) for the different coefficients U_i are thus coupled resulting in an implicit problem even for an explicit time discretization.
If, instead of the first-order advection problem, higher-order differential equations are used, more overlap integrals should be evaluated, for instance:

$$Q_{ji} = \int_{\Omega} \frac{d^2\phi_i}{dx^2}(x)\,\phi_j(x)\,dx. \tag{1.41}$$

One could opt for global basis functions or local basis functions, which are only nonzero in a limited part of the domain. The first method, called the *spectral* method, was introduced in meteorology in 1966 by, again, Robert [91] and permitted a highly accurate calculation of derivatives. In a global model spherical harmonics are the obvious choice for the basis functions, see for example the spectral IFS/ARPEGE model. For spectral NWP models defined on a limited domain, like ALADIN, double Fourier series are used as basis functions [2]. Both sets of basis functions are orthogonal resulting in a diagonal mass matrix **M**. Moreover, the differential matrices **D**, **Q**,... are diagonal too[10] and therefore the spectral method turns out to be particularly efficient in solving differential equations. For example, the Helmholtz problem Eq. (4.23), which is important for a SI NWP models, becomes a diagonal problem in spectral space.
Due to its high accuracy and efficiency, the spectral method combined with SI time discretization became very popular. Even today the combination of a spectral horizontal spatial discretization with a SI time discretization is one of the most successful approaches in NWP (e.g., IFS [116] and ALADIN [2]). Despite their success, there are some restrictions related to the use of the spectral SI method:

- The efficiency is due to the diagonal representation in spectral space of the Helmholtz problem resulting in a trivial solver, provided that there are no horizontally inhomogeneous terms included in the Helmholtz equation. Therefore, the spectral SI method is only attractive if the SI scheme used results in a constant coefficient Helmholtz problem (see 1.3.1). A more localized spatial discretization scheme (e.g., FD) is necessary if linearization around an inhomogeneous reference state is desired.
- Figure 1.5 (a) shows the first six Fourier basis functions on a small 1D domain. The global character of these functions $\phi_i(x)$ makes that all grid point values are needed to calculate one Fourier coefficient, and vice versa. This offers the high accuracy characteristic to a spectral model[11] but necessitates quite some computations and communication to go back and forth between spectral coefficients U_i

[10] except for meridional derivatives of spherical harmonics, they result in a sparse matrix
[11] One could consider the approximation of a derivative by a spectral method as a FD scheme that is using the values in all grid points of the domain.

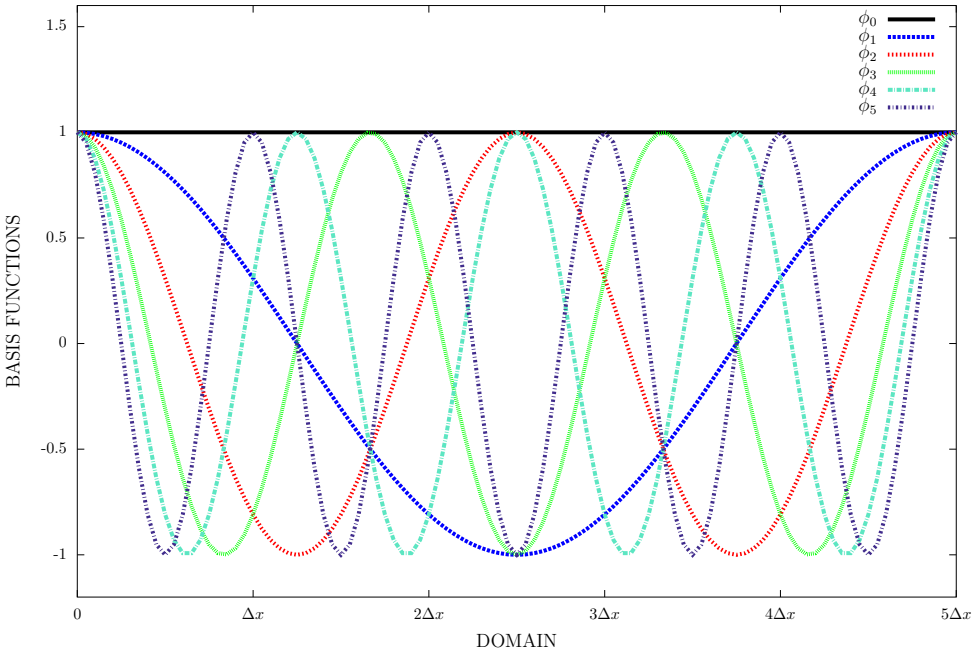
and grid point values u_i. Recently, questions were raised about the scalability on exascale computing infrastructure of these transforms.

- It is obviously impossible to construct a discontinuous function by adding up continuous Fourier (or spherical harmonics) functions. Therefore, approximations of the form of Eq. (1.33) to a field containing discontinuities will result in over and undershooting around the discontinuities, no matter how large the truncation N is chosen. This is called the *Gibbs phenomenon* and is sometimes considered as a drawback of the spectral approach [135, 64]. However, the relevance of this is questionable, since only very steep fields are affected. Furthermore, alternatives might suffer from similar problems.

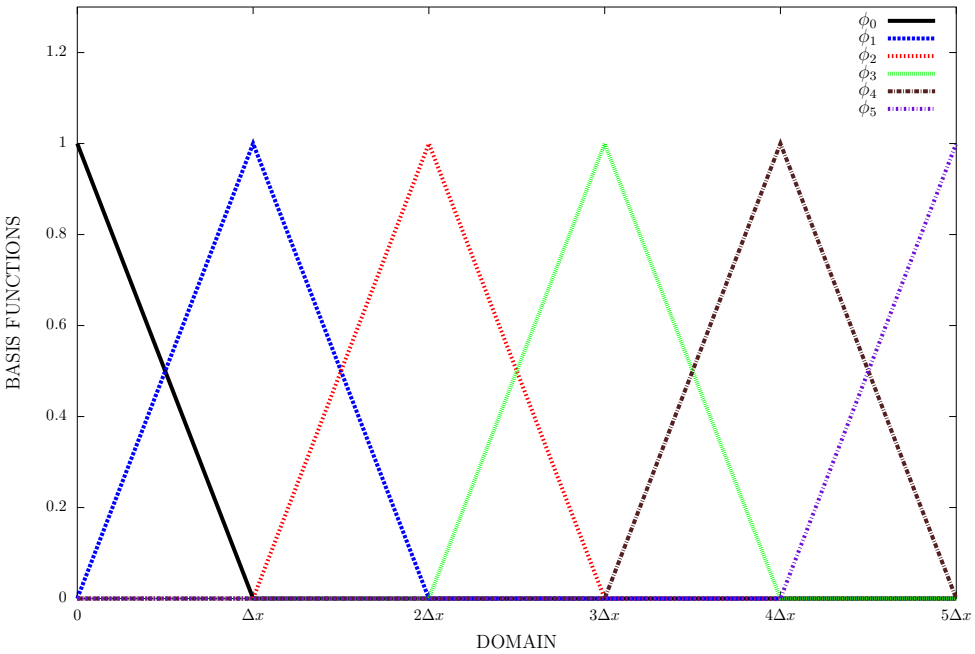
A spatial discretization scheme based on the use of local basis functions, which are only nonzero in a limited part of the domain, is called a *finite-element* (FE) method. A good introduction is found in [128], which presents the first FE scheme in NWP. FE methods are becoming more popular: although they have a lower degree of accuracy compared to spectral methods, they need less communication to calculate the expansion coefficients for given grid point values. Moreover, they can be used in a more flexible way. Figure 1.5 (b) shows the basis functions for a linear FE method, which is the most basic FE scheme. Due to the limited overlap between the local basis functions, their mass matrix $\mathbf{M}$ and differentiation matrix $\mathbf{D}$ are sparse in contrast to the sinusoidal functions (Figure 1.5 (a)) that result, counterintuitively, in diagonal matrices. In Appendix A the nonzero elements of the mass matrix $\mathbf{M}$ and differentiation matrix $\mathbf{D}$ are calculated for the linear FE method.

Many other variants of Galerkin methods exist. Two more examples are given below:

1. *Spectral elements*: this method divides the domain in nonoverlapping elements and defines high-order basis functions within each element. The method combines high-order accuracy with locality and thus limited communication. However, so-called Gaussian quadrature is used for the evaluation of integrals and due to the unequal spacing of the quadrature points there will be a very strict CFL criterion. Moreover, Melvin [70] showed that at the short scale end of the spectrum the geostrophic adjustment, an important atmospheric process explained in detail in chapter 3, is represented in an unphysical way.
2. *Mixed finite-element* method: this approach is based on the use of different sets of basis functions for different variables, for instance the basis functions for the expansion of the wind components differ from the ones for pressure. This approach can be considered as the FE variant of the staggered FD formulations. At



(a)



(b)

Figure 1.5: An expansion in global sinusoidal basis functions (a) and local linear finite-element basis functions (b).

the cost of extra complexity, this method leads to appropriate geostrophic adjustment [111, 71].

The finite-volume method

Besides FD and Galerkin methods, the *finite-volume* (FV) approach can be distinguished as a third main category of spatial discretizations. Its origins date back to 1959 [47] and the method was extended during the following decades, but it is only quite recently that it became more commonplace in NWP, mainly in US research groups (e.g., [61] developed at NCAR and NASA).

The FV method starts from the equations reformulated in their *conservative* form. In 1D this looks like

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} f(u) = S(u, x), \tag{1.42}$$

with $f(u)$ the flux determining the reaction of the flow to gradients and $S(u, x)$ a source term. By assuming a partitioning of the domain in different elements, one can spatially integrate the conservative equations over each element. For the sake of simplicity, we continue with the 1D problem. Integration of Eq. (1.42) over a 1D-element $[x_{i-1/2}, x_{i+1/2}]$ located around x_i and use of the Gauss theorem results in:

$$\frac{d}{dt}\left[\int_{x_{i-1/2}}^{x_{i+1/2}} u(x,t)\, dx\right] + f\left(u\left(x_{i+1/2}, t\right)\right) - f\left(u\left(x_{i-1/2}, t\right)\right) = \int_{x_{i-1/2}}^{x_{i+1/2}} S(u,x)\, dx. \tag{1.43}$$

Time integration from t^n to $t^{n+1} = t^n + \Delta t$ yields:

$$\begin{aligned}\int_{x_{i-1/2}}^{x_{i+1/2}} u\left(x, t^{n+1}\right) dx - \int_{x_{i-1/2}}^{x_{i+1/2}} u\left(x, t^n\right) dx + \int_{t^n}^{t^{n+1}} f\left(u\left(x_{i+1/2}, t\right)\right) dt \\ - \int_{t^n}^{t^{n+1}} f\left(u\left(x_{i-1/2}, t\right)\right) dt = \int_{t^n}^{t^{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} S(u,x)\, dx dt.\end{aligned} \tag{1.44}$$

By introducing

- the average of the u field at time t^n in element $[x_{i-1/2}, x_{i+1/2}]$

$$u_i^n = \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} u\left(x, t^n\right) dx, \tag{1.45}$$

- the average flux through the element boundary at $x_{i-1/2}$ during $[t^n, t^{n+1}]$

$$f_{i-1/2}^{n+1/2} = \frac{1}{\Delta t} \int_{t^n}^{t^{n+1}} f\left(u\left(x_{i-1/2}, t\right)\right) dt, \text{ and} \tag{1.46}$$

- the time and space averaged source term

$$S_i^{n+1/2} = \frac{1}{\Delta t \Delta x} \int_{t^n}^{t^{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} S\left(u, x\right) dx dt, \tag{1.47}$$

one obtains the following update equation

$$u_i^{n+1} = u_i^n - \frac{\Delta t}{\Delta x} \left(f_{i+1/2}^{n+1/2} - f_{i-1/2}^{n+1/2} \right) + \Delta t S_{i+1/2}^n. \tag{1.48}$$

This equation, giving the updated cell-averaged quantities, is exact. However, one needs approximations to evaluate the fluxes at the boundaries (1.46) and the averaged source term (1.47). There is many literature available on this topic, for instance, Durran [34].
We note that one of the main advantages of the FV method is its good conservation properties, as illustrated in [61]:

> The application of the finite volume algorithms for global modeling at NASA started in the late 80s and early 90s with focus on the transport process of chemical constituents and water vapor.

A recent European example of the FV method in NWP can be found in [103].

Different grids

Numerical models are formulated on a *grid* or *mesh* covering the region of interest. For limited area models this does not pose serious problems, one typically uses a rectangular Cartesian grid. However, for global modeling the grid choice is more challenging. Until now the latitude-longitude (lat-lon) grid, which defines grid points as the intersections of meridians and latitude circles, was the dominant choice for NWP. It possesses a logically rectangular structure, orthogonality, and symmetry [109], but the converging meridians result in a very high resolution over the two poles, as exemplified in the introduction of [53]. Using a reduced grid that has less grid points on latitude circles closer to the poles (e.g., the reduced Gaussian grid of [53] used for spectral models) does relax this pole problem somewhat. Due to this pole problem,

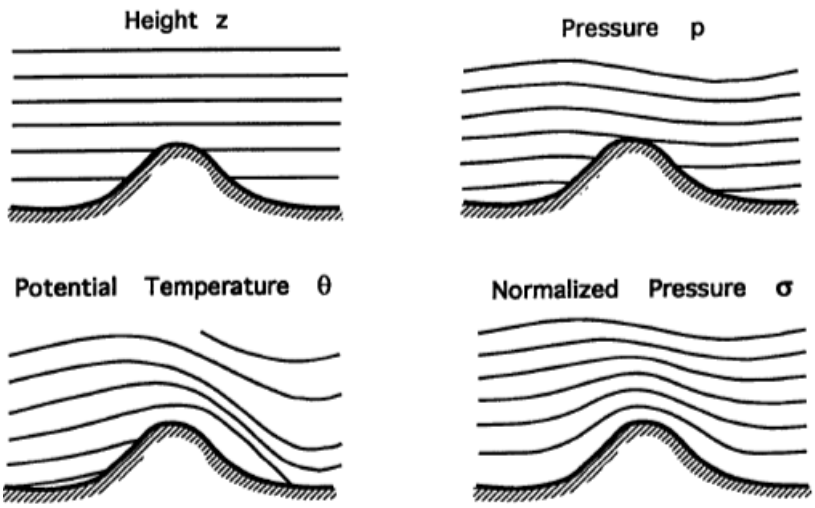


Figure 1.6: Illustration of vertical coordinate systems used in NWP. Figure is taken from [4].

the CFL stability condition bans explicit time integration on a lat-lon grid contrary to SISL discretizations, which do not suffer from this strict time step limitation. It has been advocated that also SISL schemes may become unfeasible for high-resolution lat-lon modeling on massively parallel computer architectures where communication and not computation becomes the bottleneck. This is due to the extensive data communication needed around the poles [109] which may be dealt with by a smart domain decomposition strategy as done by the IFS model [131]. The pole problem explains the renewed interest in alternative methods for global grids, as recently reviewed in [109] and [135]. Until now, the operational use of grids different from lat-lon grids is limited. Among the models listed in Table 1.1, the ICON-model, which is formulated on an icosahedral-hexagonal grid, currently is the only one.

1.3.3 Vertical space discretization

Different vertical coordinate systems exist. An overview of some of them is given in Figure 1.6. The equations have a slightly different form depending on the vertical coordinate system. For example, $\left(\frac{\nabla p}{\rho}\right)_z$, the pressure gradient term in height coordinates z, will turn into $(\nabla\phi)_p$, with ϕ the geopotential height defined as $\frac{d\phi}{dz} = g$, if pressure coordinates p are used [59]. Remark that both the height, pressure, and potential temperature coordinate surfaces cross the topography, and thus trouble the formulation of lower boundary conditions. This problem can be solved by using a *normalized* vertical coordinate. For example, the $\sigma = \frac{p}{p_s}$ (p_s surface pressure) coordinate is the normalized variant of the pressure coordinate p and this time the surface forms a coordinate surface $\sigma = 1$. However, a price is paid in the form of some extra complexity in the pressure gradient term and the continuity equation [81]. If normalized coordinates are

used, the impact of orography is not limited to the lowest levels of the atmosphere. Therefore, the coordinate surfaces higher up can differ strongly from the flow direction. To address this, hybrid vertical coordinates were introduced [97]. These coordinates vary gradually from normalized coordinates close to the surface to purely height or pressure coordinates higher up in the atmosphere.
In hydrostatic models, pressure based vertical coordinate systems are mostly used [59]. However, if the hydrostatic assumption is relaxed and the Euler equations are used, pressure is no longer decreasing monotonically when going up and therefore can be no longer used as a vertical coordinate. This explains why Euler equations were historically discretized in height coordinates until the 1992 paper of Laprise [59]. He showed that the use of hydrostatic pressure π as a vertical coordinate for an Euler system results in equations very similar to the pressure based hydrostatic equations. This result was a major breakthrough and opened the way to π-based discretizations of the Euler equations. Laprise also showed that these results could be extended to a hybrid η variant $\pi(\eta, t) = A(\eta) + B(\eta)\,\pi_s(t)$, with π_s the hybrid pressure at the surface. This coordinate system is currently used by the IFS/ARPEGE/ALADIN community.

As explained in 1.3.1, the SI approach leads to a Helmholtz equation whose operator consists of both horizontal and vertical operators. It is a serious burden to invert such a 3D problem. However, if a horizontally homogeneous reference state is assumed and if orography terms (coming from terrain following vertical coordinates) are not included in the SI operator, the Helmholtz equation can be decoupled and one can solve consecutively 1D vertical and 2D horizontal problems. However, this approach results in larger nonlinear residuals (thermal and orography) and this could hamper the stability of the scheme. Therefore, it could be advantageous to include orography and inhomogeneous temperature fields in the Helmholtz equation. The resulting Helmholtz problem is then a 3D problem that cannot be decoupled, because vertical and horizontal operators do no longer commute. For the Euler equations, deriving such a system is easier for height based coordinates (e.g., UM [28]) compared to hydrostatic pressure coordinates. This comes from the fact that in hydrostatic pressure coordinates the coordinate surfaces do evolve with time because the surface pressure $\pi_s(t)$ changes.

1.4 External constraints influence NWP choices

One interpretation of how a NWP model is constructed, was given in Figure 1.2. After presenting a wide variety of equation sets and numerical methods, an interesting question remains: which factors determine the choices made in a NWP model?

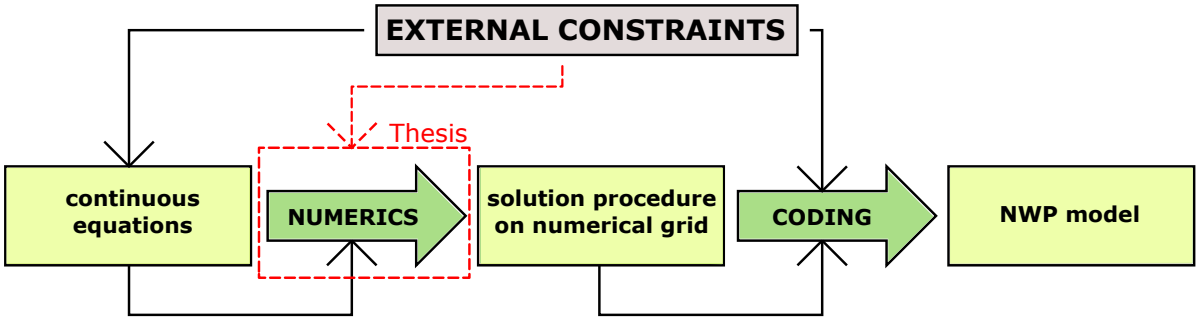


Figure 1.7: Updated schematic overview of the development of a NWP model.

An important part of the answer to this question could be gathered under the denominator *external constraints*. There is a whole series of expectations and limitations that influence or determine the model choices that are made. These constraints are often less scientific and more practical, but they are very important for model development and therefore they are added to the scheme in Figure 1.7. Remark that these constraints are external factors that are out of control of the model developer, but they will influence the equation set used, the numerics, and the code implementation. The relevant constraints for this study are listed hereafter and will be discussed in more detail in chapter 2:

- The *purpose of the model* (e.g. global climate modeling, mesoscale modeling, urban modeling,...) determines which processes should be resolved and how the domain should look like.
- There is a need for so-called *modular changes in a NWP model*. This is a practical constraint that is important in the context of new developments within an existing model framework. An operational NWP model typically has a long history, lots of different users, contributors, products, and downstream models that are coupled to the output of the NWP model,... It is therefore desirable to implement changes in a modular way that does not overhaul the complete NWP model. Changes in, for instance, the numerics should be implemented such that one could reuse the already existing code framework maximally (e.g., the time step organization of your model) and keep the coupling to downstream models and products (e.g., keep the same grid).
  It is not always obvious which path one should follow when planning developments in a NWP model. The modular methodology allows to introduce changes in an optional, gradual way with the possibility to return to the previous situation. Finally, the modular approach permits to compare different options within an otherwise unchanged model. This is interesting for a scientifically clean validation of new ideas.

- The *available high performance computing (HPC) infrastructure* forms an important constraint. First, the available computing power will determine which processes can be resolved. The computing power should be sufficient to achieve the goal of the model. Secondly, the HPC architecture will impact the efficiency and therefore applicability of different numerical schemes. Changes in supercomputing architectures may necessitate modifications in the numerics. Thirdly, the supercomputer architecture influences the computer code, for example, the parallelization strategy.

1.5 Goal and outline of the thesis

1.5.1 Goal of the thesis

To guarantee state-of-the-art models, modeling communities evaluate and improve their models continuously. This is a difficult task as the constraints for NWP models are constantly evolving. Moreover, it is hard to predict these evolutions, while the development time of new contributions to a model is often lengthy. Therefore, modeling communities should think ahead and investigate alternatives that may offer technical or scientific advantages at a certain point. This thesis is situated within such a strategy for the numerics used in the ALADIN model.

The ALADIN (**A**ire **L**imiteé **A**daptation dynamique **D**éveloppement **I**nter**N**ational) model is a limited area model (LAM, more about this in section 2.1) that is developed and used for mesoscale weather modeling and research by a consortium of 16 European and North-African countries, including Belgium [72]. Recently, the use of the model was extended to high-resolution regional climate studies [32, 46]. The ALADIN model was constructed as a limited area variant of the global IFS-ARPEGE models and therefore shares large parts of the code, for instance, the spectral SISL approach. This study is situated within and oriented towards the ALADIN framework, but conclusions are valuable for the wider NWP community.

The spectral spatial discretization methodology is generally considered as one of the most successful NWP approaches and forms the basis of the ALADIN numerics. As explained in 1.3.2, this approach coupled with a SISL time discretization combines a highly accurate horizontal spatial discretization successfully with an efficient solution procedure due to the diagonal character of the Helmholtz problem in spectral space. However, we remind that there are at least two important limitations of spectral methods:

1. Global communication, which is intrinsically linked to the spectral method, is expected to become the scalability bottleneck on future massively parallel HPC infrastructure. More details will follow in 2.3.2.
2. The spectral SISL method can only be used efficiently if a horizontally homogeneous reference state is assumed for the linearization. This may lead to large nonlinear residuals, which may form a source of instabilities.

Local horizontal spatial discretization methods, such as FD and FE methods, may have scientific as well as technical assets compared to the current spectral approach. Therefore, it is interesting to investigate local approaches within the ALADIN context. This thesis does not aim to implement different local methods in the ALADIN model or to verify the previous claims. This might be a longer term goal. This study investigates whether local methods can fit the ALADIN constraints, discussed extensively in chapter 2. What strategy should be chosen if one wants to supplement the ALADIN model with a local dynamical core? What is the price to pay if the spectral method is replaced by a local discretization scheme?

1.5.2 Outline of the thesis

This chapter sketched the general context of NWP models, with special emphasis on the equations that are used and the numerics that are employed in NWP. The presented overview is not exhaustive but prepares the reader for the study that will follow. External constraints were identified as a key factor for the choices that are made in a NWP model.

Chapter 2 lists the key constraints in the organization of the numerics of the ALADIN model. The LAM approach, time step organization, and HPC evolutions are discussed among others. These constraints define the framework wherein we can investigate local horizontal spatial discretization alternatives.

Discretization, both in space and time, influences the properties of the numerical wave solutions that are permitted in a NWP model. In chapter 3, we review the impact of the discretization choices on the dispersion properties of the numerical wave solutions present in the model. We use the shallow water equations for our review and mainly focus on the propagation of the numerical IGWs. These waves are important to restore the balance between the mass and velocity fields towards geostrophic equilibrium, one of the main equilibria in the atmosphere. By using response functions we illustrate how small changes in the numerical scheme can strongly influence the way IGWs are propagating.

Following the recommendation of the previous chapter, a detailed study of so-called Z-grid schemes is undertaken in chapter 4. We focus on Z-grid schemes that are combined with a local spatial discretization and a SISL time discretization. We present a detailed dispersion analysis and illustrate the SISL Z-grid method for toy model tests, based on the shallow water equations. At the end of the chapter, the possibility of a local method to include an inhomogeneous term in the SI problem is illustrated. This is done by solving the 1D SWE in a situation with a bottom topography that depends on the position.

Chapter 5 investigates the impact of the implementation of a local FD scheme in the ALADIN model. This is done in a minimalistic way, we mimic the use of a FD A-grid and FD Z-grid scheme by using their responses in the spectral model. This permits to diagnose some interesting issues that negatively impact the Z-grid method. The results of these model tests are discussed and can be used to identify future priorities.

Chapter 6 presents a summary of the results of this thesis and gives a prospect for future research. Finally, chapter 7 presents a concise Dutch summary of the thesis.

Constraints relevant for the numerics of the ALADIN model

2

As visualized in Figure 1.7, external constraints have a large impact on the choices that are made in the development of a NWP model. In this thesis, we study local spatial discretization schemes as an alternative for the spectral horizontal discretization used today in the ALADIN model. As new proposals should be compatible with the imposed constraints, we first discuss in this chapter those constraints that are relevant for the numerics:

1. The ALADIN consortium was founded for the purpose of delivering high-resolution forecasts over limited area domains for its member states.
2. Following the modularity constraint introduced in 1.4, changes to the ALADIN model should retain its current time step organization maximally and should still use a collocation grid, where all variables are defined in the same grid points.
3. The ALADIN model should employ the available HPC infrastructure efficiently. Projected changes in HPC architecture may favor algorithms that need little global communication. This forms one of the motivations of this thesis research.

2.1 ALADIN's aim: high-resolution modeling over limited domains

Operational weather forecasters are mainly interested in the weather for one specific region. Given a certain emission scenario, policymakers want to know the climate projections for the region, city,... they are responsible for. The previous examples illustrate that for many weather and climate applications results over only a small portion of the Earth are sufficient. Instead of running uniform resolution experiments over the whole globe, the computational resources can then be employed more efficiently by:

- making a *global forecast with a nonuniform resolution* so that the larger part of the planet is projected on a low-resolution mesh, whereas the region of interest is covered by a much finer grid that would not have been reachable with a uniform resolution model. This is done for example in the ARPEGE model, which is a global model with a so-called rotated pole, stretched grid (based on [25]) that

has its highest resolution over France. An intercomparison of variable resolution global models used for regional climate modeling is given in [42].

- making a high-resolution forecast over a small domain encompassing the region of interest. This branch of NWP is generally known as *limited area modeling*. The LAM forecast must be embedded in a lower resolution forecast over a host domain that covers the LAM domain so to provide the latter with *lateral boundary conditions* (LBCs). This is called nesting. Laprise summarized the goal of a LAM nicely in [60]:

  > The aim of LAMs is to reproduce the behavior of a global model with the same characteristics at a fraction of the computational cost.

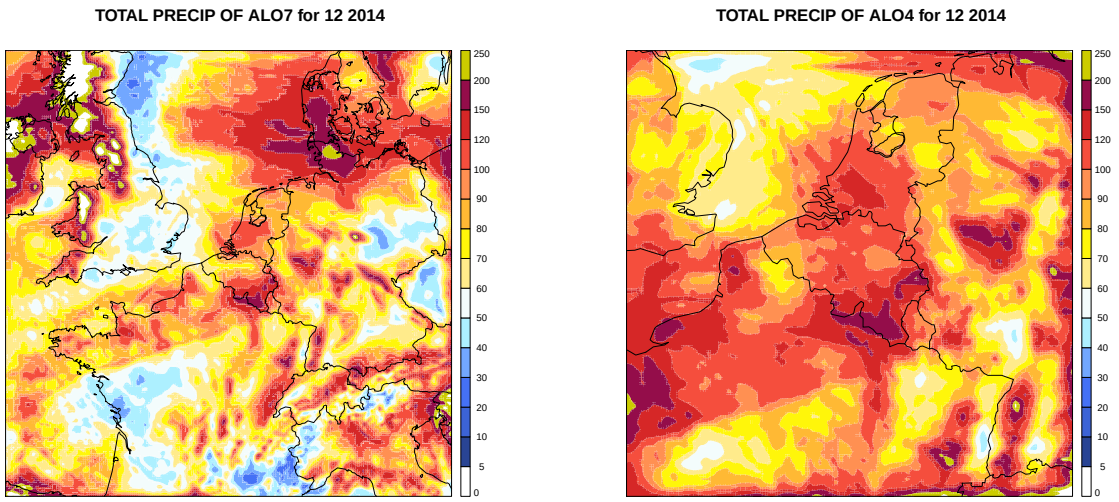
The ALADIN model, defined as a high-resolution numerical weather prediction limited area project [72], falls into the second category. We will now discuss two consequences of the LAM approach on the ALADIN numerics: the need for lateral boundary coupling (2.1.1) and biperiodization (2.1.2).

2.1.1 Lateral boundary coupling

The added value of a LAM consists of providing high-resolution meteorological details on top of the larger scale fields that are supplied by a lower resolution model. The appropriate coupling of this large scale information at the boundaries of the LAM domain is a complex task. During the forecast, the LAM will develop its own weather, which can deviate substantially from the lower resolution run used for the coupling. To avoid that applying the LBCs will result in shocks, the ALADIN model uses the so-called Davies method [26] to gradually relax the boundary values dictated by the host model to the fields of the high-resolution LAM. Symbolically this means that one should solve every time step

$$\mathbf{U} = \alpha \mathbf{U}_{\mathbf{host}} + (1 - \alpha)\, \mathbf{U}_{\mathbf{lam}}, \tag{2.1}$$

with $\mathbf{U}_{\mathbf{lam}}$ the LAM fields, $\mathbf{U}_{\mathbf{host}}$ the fields provided by the run on the host model, and $\mathbf{U}$ the resulting fields after coupling. The relaxation coefficient α will depend on the location in the domain: it will be 0 in the inner part of the domain and 1 at the boundary. In between, there is a relaxation zone where $0 < \alpha < 1$ is gradually changing. The width of the relaxation zone in ALADIN is 8 grid points. Recently, a different coupling method, called the Boyd method, was shown to deliver, at least in highly idealized tests with the ALADIN model, slightly better results [30, 119].
If LAMs are used for regional climate modeling, they are typically run on large domains for a long time. The high-resolution model then has a lot of time and space to



(a) (b)

Figure 2.1: Total precipitation for December 2014 calculated based on an intermediate resolution ($\Delta x = 7\,\text{km}$) run on a large domain (a) and a high-resolution ($\Delta x = 4\,\text{km}$) run on a small domain (b). Both LAM forecasts were coupled to the same global model. In (b) one does clearly see some unrealistic boundary patterns, which are not present in (a).

develop its own weather, which could strongly differ from the host model's solution. Some therefore argue that one should include an additional nudging of some fields over the whole domain [78].
The horizontally inhomogeneous relaxation makes that Eq. (2.1) cannot be evaluated in spectral space. The coupling is therefore done in the grid point part of the calculations in the ALADIN model.

An appropriate coupling of boundary values, supplied by a lower resolution run on a host domain, at the border of the LAM domain turns out to be crucial if one wants to create added value by the LAM forecast. If not done properly, the LAM forecast will show unphysical behavior originating from the boundaries. Warner [129] listed conditions that should be met to obtain appropriate coupling. Apart from limiting the resolution difference between the host model and the LAM, having a decent coupling update frequency ([118] for an analysis), and using similar physics, the way to apply the LBCs is identified as very important. Even if done carefully, effects related to the boundaries cannot be completely avoided. As an illustration, consider Figure 2.1 where the monthly accumulation of precipitation is shown for December 2014 as predicted by the operationally used ALADIN runs. In Figure 2.1 (a) the precipitation is calculated for 7 km resolution runs on a large domain, whereas Figure 2.1 (b) is based

on 4 km runs on a smaller domain[1]. Both runs were coupled to the same global model. As expected, the two maps largely show the same pattern with some additional high-resolution features over Belgium in Figure 2.1 (b). However, in the boundary region of the high-resolution domain large amounts of precipitation that are not observed in the low-resolution one appear. This is clearly due to the lateral boundary coupling. Coupling the 4 km run to the fields obtained with the 7 km forecast instead of the global model may improve the quality of the 4 km run seriously [129, 29].

2.1.2 Biperiodization

At first sight, a spectral horizontal discretization seems incompatible with a LAM. The basis functions used are bi-Fourier functions that are periodic along the x- and y-direction. However, the meteorological fields that should be represented as a linear combination of these basis functions are not periodic. An engineering trick is used to solve this problem. Along the domain[2] some extra rows and columns of grid points are added. This so-called extension zone contains for each field values interpolated between the fields at the boundary of the physical domain for this row or column. In this way, including these nonphysical values in the extension zone results in bi-periodic fields, which are suitable for a decomposition in bi-Fourier functions.
In ALADIN the biperiodization procedure is undertaken, before starting the LAM run, on the fields coming from the host model that will be coupled to the LAM model. Due to the relaxation factor $\alpha = 1$ at the boundaries, the coupling procedure (2.1) results in biperiodic fields. More detailed information can be found in [119].

2.2 Modularity constraints on numerics for ALADIN

The need for modular changes in a NWP model was identified as a second constraint in 1.4. Apart from its practical and organizational grounds, there is also a scientific motivation for modularity as it permits clean comparisons of different implementations.
For this thesis, we impose that two methodologies, which are central to the current spectral ALADIN model, need to be conserved as closely as possible: the organization of the model computations during one time step and the use of a collocation grid. To clarify the time step organization, we will first come back on the notion of *physics parameterizations*, which was shortly touched in chapter 1.

[1] The data used are the output between +6 h and +12 h of the operationally used 4 km and 7 km runs of the RMI.
[2] In ALADIN at the top and eastern boundary.

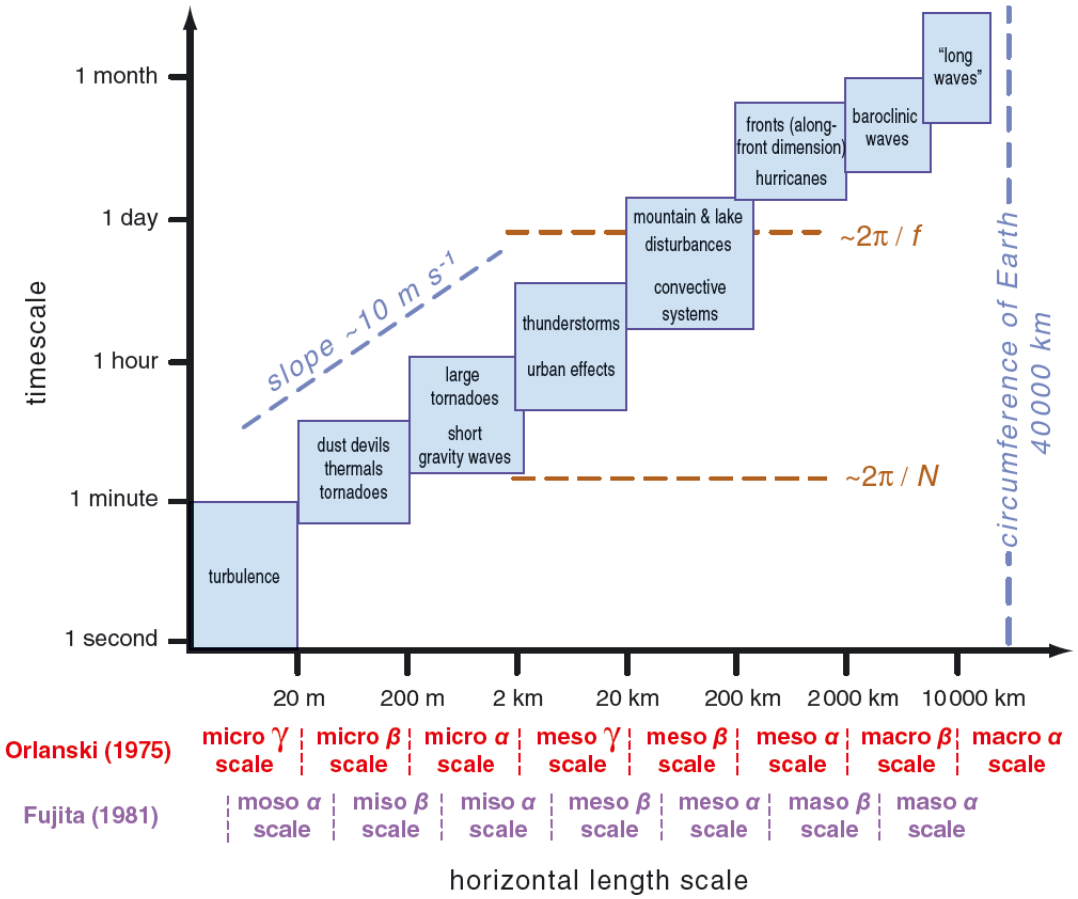


Figure 2.2: Overview of the spatial and temporal scales of atmospheric motions, f stands for the Coriolis constant and N for the Brunt-Väissälä frequency (see 3.1). This figure is taken from [66].

2.2.1 Physics parameterizations of unresolved processes

Section 1.2 presented a review of the fluid dynamics equations that are used to describe motions in the atmosphere. After studying some filtering approximations, it was stated that the evolution of the atmosphere, represented by the prognostic state vector $\mathbf{U}(t)$, could be described symbolically by

$$\frac{\partial \mathbf{U}(t)}{\partial t} = \mathbf{A}(\mathbf{U}(t)) + \mathbf{M}(\mathbf{U}(t)) + \mathbf{P}(\mathbf{U}(t)), \tag{2.2}$$

with $\mathbf{A}$ representing advection, $\mathbf{M}$ resolved forcings, and $\mathbf{P}$ tendencies coming from unresolved processes. Before having a closer look at the contribution of the unresolved processes, we should first introduce the typical spatial and temporal scales met in meteorology.

Figure 2.2 presents an overview of the scales of meteorological phenomena. It is striking that larger motions are also longer living. At the short scale end, with dimensions

in the order of seconds and meters, there is turbulence whereas the long waves, also known as Rossby waves, form the other extreme with their planetary scales and life expectancy of months. Atmospheric motions are typically subdivided in different categories whose names and limits depend on the author, as illustrated below Figure 2.2. Without specifying the precise limits of each category, this thesis will use the terms microscale, mesoscale, synoptic scale (short end of macroscale), and planetary scale (long end of macroscale).

Due to the limited computational resources and the large spectrum of spatial and temporal scales at which physical processes that are relevant for meteorology are taking place, it is impossible to resolve all of them. Processes that are not explicitly resolved by the model resolution or time step are accounted for by calculating their average impact on the time evolution $\frac{\partial \mathbf{U}}{\partial t}$ of the prognostic variables. This methodology is called parameterization. The impact of an unresolved process will depend on the atmospheric state. Paramaterization schemes differ in complexity and some need vertical calculations in the air column above a grid point. A typical meteorological model counts many parameterization schemes. In the ALADIN model all the combined parameterization tendencies are collected into one physics tendency $\mathbf{P}(\mathbf{U}(t))$ and added to the prognostic equation (2.2). It is obvious that the calculations of the physics parameterization, just like the LBC coupling, needs to be done in grid point space.
Radiation, cloud microphysics, land-surface interaction, deep convection and turbulence are some of the most important subgrid physical processes that need parameterization. There is an important difference between the first three and last two processes. The first three are described by laws (e.g., radiative transfer equations) that have no relation with the Euler equations. Independent of the resolution used, they should always be parameterized. However, deep convection and turbulence are described by the Euler equations as soon as the model grid is sufficiently fine. Table 2.1 presents an overview of some important atmospheric motions, their corresponding horizontal scales, and whether they are resolved by a current day global, LAM, or large eddy simulation (LES) model. For example, deep convection characterized by the development of cumulonimbus clouds, is currently unresolved in global models ($\Delta x \sim 16\,\text{km}$) but at least partly resolved in LAMs ($\Delta x \sim 2.5\,\text{km}$). Most NWP communities therefore gradually switch off their deep convection parametrization scheme as they enter the *gray zone* for convection ($\Delta x < 10\,\text{km}$). Developing a scheme that couples coherently resolved and unresolved features (to prevent double counting) and thus works consistently for different resolutions seems a good practice to approach such gray zone issues [45]. Table 2.1 illustrates an important reality: LAMs explore new scales and their difficulties before global models and therefore the latter can learn from the strategy used

Table 2.1: Overview of the different scales of atmospheric motions based on Table 1.4 of *An introduction to Dynamic Meteorology* of Holton [50]. The three rightmost columns show which phenomena are resolved by a current day global model, LAM, and large eddy simulation (LES) model.

Type of motion	Horizontal scale (m)	Global model	LAM	LES
planetary waves	10^7	✓	×	×
mid-latitude cyclones	10^6	✓	×	×
hurricanes	10^5	✓	✓	×
fronts	$10^4 - 10^5$	✓	✓	×
cumulonimbus clouds	10^3	×	✓	✓
tornadoes	10^2	×	×	✓
gusts	$10 - 10^2$	×	×	✓
turbulent eddies	$1 - 10$	×	×	✓ or ×

by the LAMs. While LAMs are leaving the gray zone of convection, they are running into a *terra incognita* for turbulence, as explained in [138]. And here LAMs could learn from the LES models, which are now already partly resolving turbulence.
Despite having a sufficiently fine resolution, some model types cannot be used to simulate large scale phenomena because their typical domain size is too small to contain these horizontal scales. As an example Table 2.1 shows that a LES model or a LAM is not able to resolve a planetary wave. This explains why often multiple models with different resolutions are nested e.g., to model the meteorological influence of gas dispersion in an urban district [73] or to make regional climate projections [46].
Finally, one must realize that the effective resolution of the model is not identical to the grid distance Δx of the model. In [100] the effective resolution is defined as *the wavelength where a model's spectrum begins to decay relative to the observed spectrum or relative to a spectrum from a higher resolution simulation*. The reasons for this deviation of model spectra for the shortest waves are twofold. First, the shortest scale phenomena are often poorly represented by the numerical integration scheme used and secondly, diffusion (both explicit and implicit due to numerics) or truncation is applied at the short end of the wave spectrum. Studies to determine effective resolutions are based on the definition given above: one compares the spectrum of the model output with observational spectra or spectra obtained with finer resolution models. In [100] an effective resolution around $7\Delta x$ was found for WRF NWP simulations, whereas [1] found $8\Delta x$ as effective resolution of the IFS model on a linear grid. More recent experiments with IFS on a cubic grid indicate that this approach may offer a higher effective resolution [65].

Table 2.2: The different steps of interest carried out during one time step integration of the ALADIN model. The equivalent of each term in (2.3) is given in the second column of the table (based on Table 2 in [117]).

	ALADIN time step organization	
1	transform fields: spectral $\rightarrow$ grid point	
2	calculate physics in arrival points	$\mathcal{P}(\mathbf{U}_A^0)$
3	update tendencies	
4	compute SL departure points D and do interpolations	
5	compute explicit part dynamics	$(\mathcal{I} + \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{U}_D^0 + \Delta t(\mathcal{M} - \mathcal{L}^*)\left(\tilde{\mathbf{U}}\right)$
6	add all tendencies	$\mathbf{R}_{lam}$
7	lateral boundary coupling	$\mathbf{R}_{tot} = \alpha\mathbf{R}_{host} + (1-\alpha)\mathbf{R}_{lam}$
8	transform fields: grid point $\rightarrow$ spectral	
9	solve for updated fields	$\mathbf{U}_A^+ = (\mathcal{I} - \frac{\Delta t}{2}\mathcal{L}^*)^{-1}\mathbf{R}_{tot}$

2.2.2 Time step organization

In this subsection the way the ALADIN model calculations are organized during one time step is reviewed. Alternative discretization schemes should maximally obey the current time step organization.

Let us use the prototype equation (2.2) to outline the different steps of the ALADIN time step organization. The ALADIN model is based on a 2TL SISL discretization scheme. Applying this approach to Eq. (2.2) yields

$$\begin{aligned}(\mathcal{I} - \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{U}_A^+ &= (\mathcal{I} + \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{U}_D^0 + \Delta t(\mathcal{M} - \mathcal{L}^*)\left(\tilde{\mathbf{U}}\right) + \Delta t\mathcal{P}(\mathbf{U}^0) \\ &= \mathbf{R}, \end{aligned} \tag{2.3}$$

with $\mathcal{I}$ is the identity operator, Δt the time step, $\mathbf{U}$ the vector containing the prognostic variables in all the grid points, $\mathcal{M}$ the complete dynamics, $\mathcal{L}^*$ the dynamics linearized around a horizontally homogeneous reference state, and $\mathcal{P}$ the physics. The spatial discretization is included in the operators $\mathcal{L}^*$, $\mathcal{M}$, and $\mathcal{P}$. The superscripts $^+$ and ^{0} represent evaluations at time $t + \Delta t$ and t, respectively. The subscripts $_D$ and $_A$ imply that the fields should be evaluated in the departure points and the arrival points, which coincide with the grid. Remark that, despite being second-order accurate for the linear part of the dynamics $\mathcal{L}^*$, the overall accuracy is only first-order.

Table 2.2 shows the ALADIN procedure to solve Eq. (2.3). The tendencies of the physics $\mathcal{P}(\mathbf{U}^0)$ are first calculated in the arrival points (step 2), but they are later interpolated to the departure points (step 4) where they are added to the tendencies of the explicit

dynamics (step 6). The different contributions to the physics are calculated independently and aggregated to one tendency [117]. For the ALADIN experiments that will be presented in chapter 5 a set of physics parameterizations known under the name ALARO-0 is used. Information about the parameterization strategies used in this package can be found in [32] and [46].
Different options exist for the evaluation of the nonlinear part of the dynamics $(\mathcal{M} - \mathcal{L}^*)(\tilde{\mathbf{U}})$, among others the *stable extrapolation two-time level scheme* (SETTLS) described in [52], which enhances stability while being second-order accurate (step 5). The coupling to the host model, which was explained in 2.1, is done in grid point space on the right hand sides (details described by Radnoti [87]) by relaxing the LAM fields to the fields resulting from a run on a larger host domain (step 7). Due to the periodicity of the coupling fields, this immediately leads to periodic right hand sides, ready to be transformed to spectral space (step 8). An overview of the biperiodization and coupling methods used in ALADIN can be found in [119].

2.2.3 Collocation grid

The ALADIN model is defined on a so-called collocation or unstaggered grid [68]. On such a horizontal grid all variables are defined in the same grid points. For reasons that will become clear in chapter 3, it may be advantageous to combine local spatial discretization methods with a staggered grid formulation where the pressure and wind velocity variables are evaluated in different grid points. However, leaving the collocation grid would imply that the post-processing, physics parameterization, and SL calculations should be reformulated. It would also treble the amount of trajectory calculations. Maintaining a collocation grid is therefore an important constraint.

2.3 High performance computing as an evolving constraint

2.3.1 Introduction

As realized already by Richardson [85], a powerful computing system is needed to forecast the weather *on time*. The first numerical weather forecast, which was performed by Charney, Fjörtoft, and Von Neumann, is described in the 1950 Tellus article *Numerical Integration of the Barotropic Vorticity Equation* [18]. This meteorological breakthrough was achieved on the ENIAC (Electronic Numerical Integrator And Computer) built during World War II, which was the first general purpose machine and a cornerstone in the development of and problem solving with computers. Apart from military calculations, the ENIAC was available to researchers to solve problems in different do-

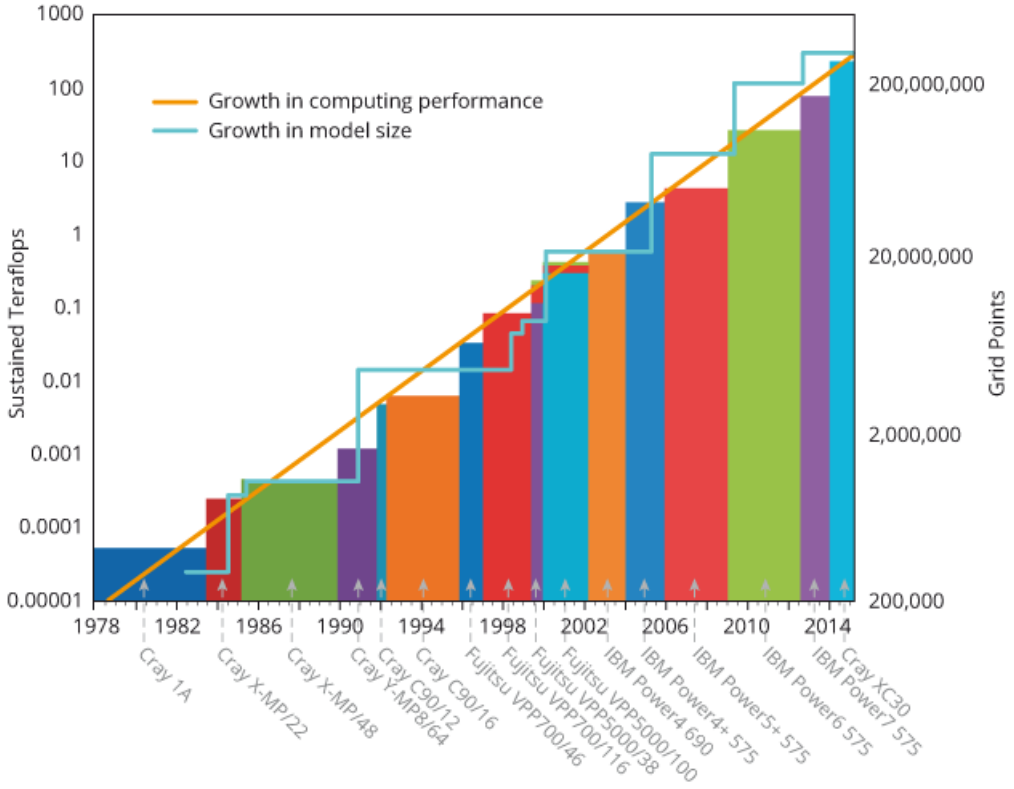


Figure 2.3: Evolution of sustained performance of the HPC infrastructure of ECMWF (expressed in teraflops, left axis) and the number of grid points used for their global model run (right axis). This figure is taken from [49].

mains and weather forecasting was among them [43]. Despite its huge (at least for that era) computing power, the ENIAC needed about 24 hours to perform a 24 hour integration of the barotropic vorticity equations on a 15 by 18 grid points LAM domain. Although still lightyears away from the operational NWP standards as we know them today, this test was of large importance as it demonstrated the possibility of weather forecasting by numerical integrations.

Ever since, the computing capacities have increased dramatically paving the way towards integrating more complete and thus more complex sets of equations at higher resolutions. Figure 2.3 illustrates this increase in computing power for the European Center for Medium-Range Weather Forecasts (ECMWF) from the seventies till today. ECMWF is considered as the reference for medium range weather forecasting and its computing facilities have always been cutting-edge. It thus forms a good probe to study trends in supercomputing in NWP. The horizontal axis forms a timeline and on the left vertical axis the computing power available for operational weather forecasting at that time is plotted. Computing power can be expressed by different quantities, here *sustained performance* expressed in teraflops ($1\,\text{Tflops} = 10^{12}\,\text{flops}$) is used. One

flop (floating point operations per second) corresponds to one arithmetic operation (addition or multiplication) per second. Sustained performance reflects the number of computations per second attained for a given application (in Table 2.3 a benchmark test representative for the operational IFS forecast) and is often a better measure than the *peak performance*, which is the theoretical maximal performance of the machine. This implies that the sustained performance is always lower than the peak performance. As an example, the two most recent ECMWF supercomputers had a sustained performance that was respectively 75 and 18 times less than their peak performance. The same can be seen in the top500 list (http://www.top500.org/), which publishes both the LINPACK sustained performance and the peak performance of the 500 most powerful supercomputers in the world. However, in this list the sustained and theoretical performance are much closer to each other. This illustrates the complexity of the NWP problem and it shows that a code organization tuned to the computational infrastructure is crucial to get the maximum out of your machine.
Figure 2.3 also shows the increase in resolution by plotting the number of grid points ($N_{tot} = N_x N_y N_z$ with N_x, N_y, and N_z the number of grid points in the three directions) of the used grid. The resolution increased substantially but not at the rate the computing power has been growing.

To appreciate the speed at which computing power is evolving, compare the performance of a typical smartphone today with the evolution of the ECMWF computing power. The Samsung Galaxy S6 has a peak performance of 34.8 GFlops. To make a fair comparison, let us assume that the sustained performance is 20 times lower than the theoretical performance. It can be seen that the computing capabilities of such a smartphone beat the ECMWF supercomputers of the early nineties.
The new cluster used for the operational weather forecasts for Belgium, in use since the end of 2015, has a peak performance of 40 TFlops. By using a LAM, this limited computing power can still be employed for high-resolution weather forecasts.

As the electricity costs are becoming very large for NWP centres there is recently increased awareness about energy-to-solution e.g. [131]. This quantity will partly depend on the technology used in the HPC system, but also partly on the algorithms used. This topic falls out of the scope of this thesis.

In the next section we will have a more detailed look at the relation between the numerical algorithms used and computer architecture which is very important for the time-to-solution.

2.3.2 Link between HPC architecture and numerics

Apart from the computing power, the architecture of supercomputers did evolve. In the early days of supercomputing the computations were done by one computing core. Computer scientists managed to drastically improve the computational power of cores. However, to accomplish the needs of researchers and business, machines consisting of multiple cores were developed. Today, a supercomputer consists of thousands of computing units (the newest Cray XC30 at ECMWF consists of 84120 cores) and this trend towards massively parallel computing is expected to continue. Today's most powerful machine, Tianhe 2, has over three million cores. It is of large importance that not only the computations but also the memory is distributed over many components. As motivated earlier, the numerical problem to solve must take into account the architecture of the machine (or the other way around but it is clear that NWP is not steering the supercomputer trends). An early day 'supercomputer' example to illustrate the impact of technological innovations on the algorithm used [43]:

> Solving problems using the ENIAC often required the development of new methods in numerical analysis, since numerical methods previously used were based on the assumption of slow computing (as then available) but essentially unlimited storage (paper had been plentiful for years).

The current trend of thousands of cores connected by a communication network is clearly favoring methods that are computationally intensive but need only limited communication. Model discretizations that do need lots of communication could end up in a situation where the processors are often idle because they have to wait on data that are dragged by the (relatively) slow communication network.
Cats undertook an interesting exercise, described in an unpublished note written in 2008 [17]. He estimated the impact of the evolving HPC technology on two different discretization approaches for the dynamical core:

- Type 1, the explicit Eulerian grid point model. Apart from being a HEVI scheme (and thus semi-implicit in the vertical), ICON [140] suits more or less this definition.
- Type 2, the SISL spectral model. The IFS/ARPEGE/ALADIN [2] family is an example of this approach.

Given some assumptions, Cats undertook a quantitative study to estimate the performance of the two model types on future supercomputers. In what follows a qualitative discussion will be presented.
Type 1 models need short time steps due to the CFL stability criterion and they will therefore need many time integrations. However, the local grid point calculations (e.g.

FD) will limit the need for communications. The calculation of derivatives is only based on the values in the neighboring grid points and the higher the order of the used discretization, the larger the stencil of the involved grid points becomes. These methods scale well because the corresponding communication between cores is *local* if the computer topology reflects the grid structure. Finally, the explicit time integration results in a direct update of the prognostic variables.
The SISL spectral models are unconditionally stable (at least for the linearized part of the equations) and are therefore used with long time steps. Type 2 models will thus need far less integrations than type 1 models. This advantage is partly offset because an integration is more expensive. The calculation of the derivatives with spectral accuracy is straightforward once the spectral coefficients are known. However, as illustrated for the ALADIN model in Table 2.2, part of the computations, such as physics parameterizations, SL advection, and LBC coupling, should be done in grid point space. The transformations from grid point to spectral space form the main challenge of the type 2 models. We will have a more detailed look into these transformations.
Spectral methods are typically based on sinusoidal basis functions $\phi_i(x,y)$[3]. The transformations between grid point and spectral space of a function $u(x)$ sampled in N equally spaced points over the 1D domain $[0, L]$ read as

$$U_k = \sum_{j=0}^{N-1} u_j e^{-ik(j\Delta x)\frac{2\pi}{L}},\ \forall k = 0,1,...,N-1 \quad (\text{grid point} \longrightarrow \text{spectral}) \text{ and} \tag{2.4}$$

$$u_j = \frac{1}{N}\sum_{k=0}^{N-1} U_k e^{ik(j\Delta x)\frac{2\pi}{L}},\ \forall j = 0,1,...,N-1 \quad (\text{spectral} \longrightarrow \text{grid point}), \tag{2.5}$$

with U_k the spectral coefficient corresponding to the wave with wavenumber k and thus wavelength $\frac{L}{k}$. Equations (2.4) and (2.5) have two important drawbacks:

1. They scale quadratically (the number of operations is $\sim O\left(N^2\right)$) in function of the number of grid points N. To calculate the N spectral coefficients one does, indeed, need N multiplications and $N-1$ additions for each spectral coefficient thus $N \times (2N-1)$ operations. This is a weakness compared with FD derivative calculations, which scale linearly ($\sim O(N)$). However, there exists an algorithm, the so-called Cooley-Tukey Fast Fourier Transforms method, that evaluates the equation in a $\sim O(N\log N)$ way. Variants of these method are used in NWP and this seems to make the computational problem less stringent. For a more detailed analysis of the number of operations used for the transformations in NWP models the reader is referred to [67]

[3] This is true for the spectral LAMs. However, global methods will use spherical harmonics, consisting of a wave in the longitudinal direction modulated by an associated Legendre polynomial in the meridional direction.

2. The high-order accuracy of the spectral derivatives is due to the global character of the method: all grid point values are used to calculate a derivative in one point. This global character is reflected in the equations (2.4) and (2.5) where e.g., all grid point values are used to calculate one spectral coefficient U_k. If one would evaluate Eq. (2.4) starting from data organized in a domain decomposition way where each core contains the data of one horizontal region of the domain, lots of communication would be needed during the computations. It is important to organize the communication in a well adapted way. The *transposition method*, used for example in the IFS/ARPEGE/ALADIN family of models, reorganizes the data before evaluating Eq. (2.4) and (2.5) such that every core has already all data it needs for the calculations. In this way, there is no need for communication during the computational phases. The transposition strategy is illustrated in Figure 2.4. Starting from a horizontal domain decomposition data distribution over the cores (for the physics, explicit dynamics,...), the data are redistributed towards a set-up where all data for one latitude are collected onto one core. This permits an efficient on-core evaluation of the Fourier transforms in the longitudinal direction. Then the data are redistributed so that data with the same longitudinal wavenumber m are on the same core permitting an easy evaluation of the meridional Fourier transforms. Similar transpositions are needed to go back from spectral to grid point space [37].

The current trend towards massively parallel supercomputers makes the communication needed for the transposition far more challenging than the first, computational scaling drawback. The *global* communication needed for the transpositions does not scale, in contrast with the earlier mentioned local communication. It is therefore considered as a limitation to the use of spectral methods, as mentioned in most papers describing models not using the spectral method e.g., [140]. Also Cats concluded that one day the spectral SISL method would be outperformed by the local explicit Eulerian models [17]. This cross-over point could shift if the technological HPC evolutions would undergo serious changes. Cats explained that using grid computing for HPC (computers are then located far away from each other) would accelerate the spectral redemption [17]. Another technological game changer could be the use of silicon photonics so that communication could take place at much higher speeds [94]. Some research groups are currently developing optical devices (e.g., switches, interconnects,...) to enable the photonics to take over functions done now by the slower electronics. Such a technological revolution could prolong the use of communication intensive algorithms.

However, it is dangerous to make strong claims about future evolutions. It is, for

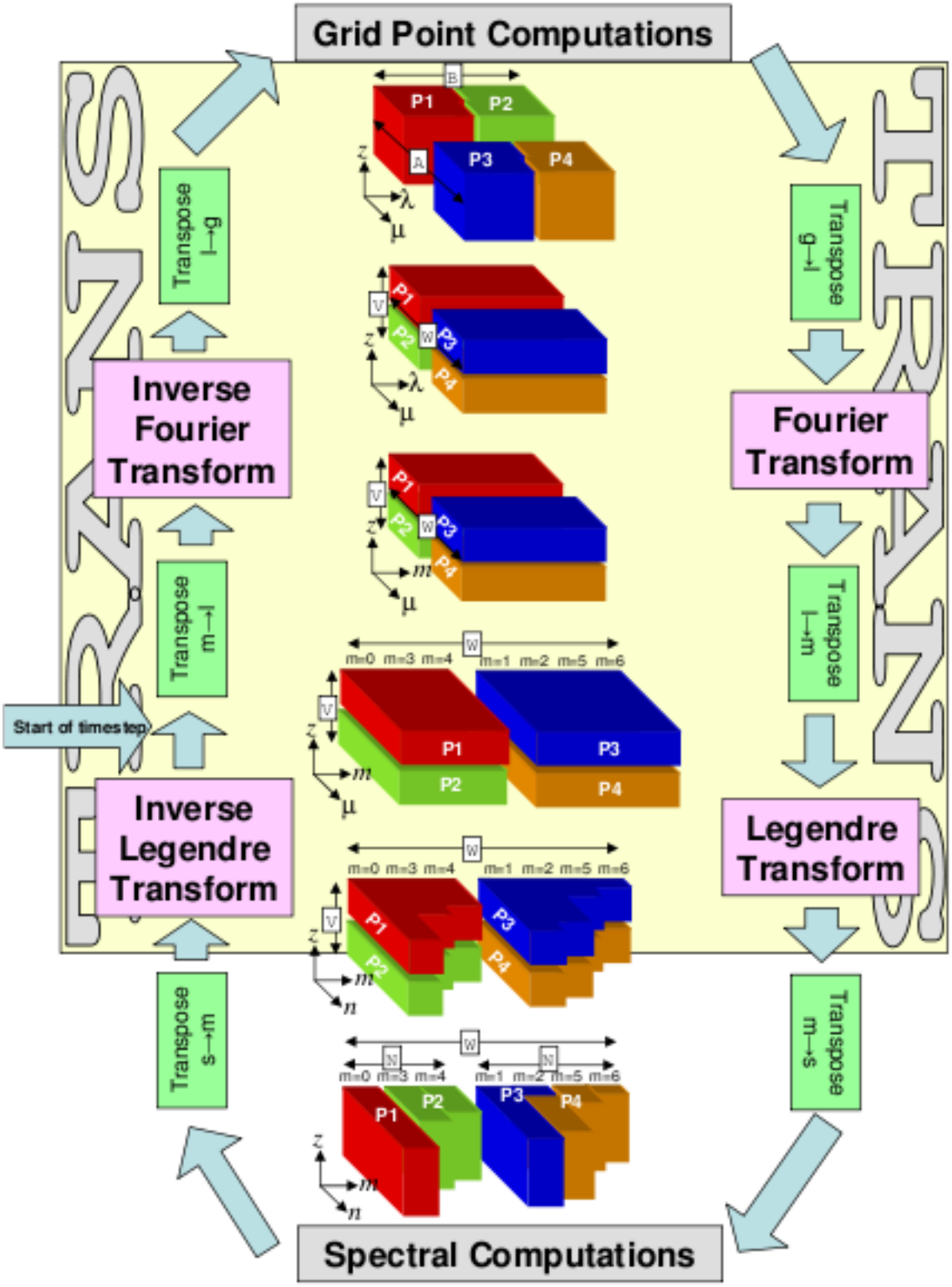


Figure 2.4: Transposition strategy of the spectral IFS code for four computing cores. The figure illustrates the different transpositions (and thus communication phases) needed to shuttle between grid point and spectral space. The ALADIN model has a very similar procedure, except that the bi-Fourier functions need twice a Fourier transform instead of the Fourier and Legendre transform needed for the spherical harmonics. Figure from [37].

example, not the first time the spectral SISL method is redeemed. In 1990 Côté and Staniforth expected the spectral method to be outperformed soon due to the computationally expensive Legendre transforms (only in global models) [20]:

> It is believed that the speed of present-day computers is such that the crossover point where finite-difference/finite-element methods become more cost-effective than the spectral method has now been reached for medium range forecasting.

However, during the consequent year the computational burdens of the Legendre transforms were weakened by a reduction of the number of transforms by reformulating the algorithm [114] and by using reduced Gaussian grids, which have significantly less grid points [53]. So Côté wrote two years later [21]:

> In the meantime, Hortal and Simmons and Temperton have shown, respectively, that some economy may be obtained by reducing the number of collocation points and transforms, and that we were perhaps unduly pessimistic for the immediate future.

And now, more than 20 years later, the spectral method is still used for medium range forecasting. However, the communication bottleneck, which is the current day threat for spectral, is linked fundamentally to the philosophy of the spectral method. Miraculous solutions do not exist for this problem. That is why spectral NWP communities are actively studying this topic today. They want to ensure that their NWP models will use future exascale machines efficiently.

2.4 Conclusion

Practical constraints were identified as a decisive factor in the development of an operational atmospheric model. This chapter highlights three constraints that have a large impact on the ALADIN numerics:

1. The ALADIN model was established as a high-resolution LAM model implying the need for LBC coupling.
2. Given the multitude of applications and products linked to the model and the complexity of the code, one needs to implement changes in a modular way to minimize the impact on the existing code. Relevant for the spectral discretization are the time step organization and the use of a collocation grid.
3. There is a clear evolution in HPC architecture towards massively parallel machines. This may create scalability problems for algorithms that need lots of

global communication e.g., the Fourier transforms. This trend in HPC is one of the motivations to rethink the horizontal spatial discretization of ALADIN.

In the following chapters we take these external constraints into account, while investigating local alternatives for the current spectral spatial discretization.

Impact of discretization on numerical representation of atmospheric waves

3

The equations governing the motions in the atmosphere do support many different wave types. Some of them have a direct impact on the weather, others are of no importance to meteorology. The assumed approximations to the Euler equations, the chosen space and time discretization, and the chosen grid determine the properties of the numerical representation of the atmospheric waves. An appropriate atmospheric model should result in a realistic numerical representation of the relevant atmospheric waves.

Due to the exact calculation of derivatives in spectral space, the spectral method for the horizontal spatial discretization conserves the analytical behavior of the numerical waves. No surprise that plugging in a local spatial discretization method will degrade the numerical representation of waves. One should therefore study the impact of the local alternatives on the wave behavior and decide whether the unavoidable distortions are acceptable. In this chapter a review and discussion about numerical wave behavior is presented.

Section 3.1 presents an overview of the spectrum of atmospheric waves. It explains why studying the strongly simplified shallow water equations (SWE) is a good way to reveal numerical wave distortions. The impact of time and space discretization, with focus on the arrangement of the variables on the grid, is reviewed in section 3.2 and 3.3. Finally, the consequences of these findings for our thesis are discussed.

3.1 Atmospheric waves and the shallow water equations

In this section we first introduce the main wave types present in the atmosphere. Secondly, we discuss how and why the SWE can be used to test the impact of discretization schemes on the numerical waves.

3.1.1 Types of atmospheric waves

Our atmosphere supports different types of waves. Assume, for example, a dry and stably stratified atmosphere, which corresponds to $\frac{d\theta}{dz} > 0$ wherein θ represents the

potential temperature of the atmosphere and z the height. An air parcel initially at altitude z_0 that is displaced adiabatically upwards over a certain distance $\delta z > 0$ in a statically stable atmosphere ends up in an environment that has a higher temperature than the parcel. This is because $\theta(z_0 + \delta z) > \theta_{par} = \theta(z_0)$ with θ_{par} the potential temperature of the air parcel. As a consequence, buoyancy forces the particle down and at the starting height z_0 the particle feels no more buoyancy implying no acceleration at that point. Consequently, the air parcel reaches at this point its maximal downward velocity. Once below its initial height z_0, the air parcel has a temperature higher than its environment. Now buoyancy works upward resulting in an upward acceleration. The downward velocity decreases, becomes 0 at its minimal height, and from this point on the parcel undergoes, again, an upward motion. In this way, the air parcel that was moved away in a stably stratified atmosphere from its initial, equilibrium height undergoes a *buoyancy oscillation*. However, the oscillation will not be limited to one parcel. Neighboring parcels are influenced and start to oscillate too (but with a certain delay or phase change). We now have a so-called *gravity wave* with parcel oscillations perpendicular to the propagation direction of the wave.
Different wave types are included in the Euler equations:

- *gravity waves*, caused by buoyancy oscillations, as explained above. The static stability of the atmosphere controls the frequency of the buoyancy oscillations $N = \sqrt{g\frac{dln\theta}{dz}}$ called the Brunt-Väissälä frequency. Mountain waves excited when statically stable air flows over orography are one example of atmospheric gravity waves. In a statically unstable atmosphere ($\frac{d\theta}{dz} < 0$) this frequency will be an imaginary number. Displacing a parcel will no longer result in an oscillation, parcels will keep rising and this may be the onset of convection.

- *inertial waves*. These waves appear when in a situation of geostrophic balance an air parcel is, for some reason, moved away over a certain distance. The parcel then has a velocity different from the geostrophic flow and the Coriolis force will therefore act on this particle trying to restore the balance. An important difference compared with the buoyancy oscillations consists of the direction of the restoring force. The buoyancy force is always acting along the same direction, the vertical, whereas the Coriolis force will point in a nonconstant direction perpendicular to the moving parcel. That is why a parcel brought out of geostrophic balance will travel along elliptical trajectories. These inertia oscillations are called inertia waves.

- *inertia-gravity waves or inertio-gravity waves (IGWs)*. In the atmosphere the two previous restoring forces can work together on an air parcel resulting in a so-called IGW. As an example, consider a buoyancy oscillation with horizontal scales suf-

ficiently large such that it is also influenced by the Coriolis force. In the limit of purely vertical or horizontal oscillations this will correspond to pure gravity and inertial waves. The direct importance of IGWs seems to be limited to mesoscale phenomena. They are, for example, thought to play a role in clear air turbulence, which poses an annoying problem for civil aviation [127]. Still today, the role and importance of IGWs remains unclear. Observational studies are undertaken to trace IGWs in the atmosphere and learn more about their properties, behavior, origins,... In [126], for example, a radar investigation to study IGWs in the troposphere and lower stratosphere is described. In their introduction they list potential sources for IGW generation:

> There are several theories in the literature for the generation of IGWs, e.g. diurnal variation in convection, instability of a horizontal shear line, instability of a vertical shear layer or flow over mountains. The most popular is probably that of 'spontaneous' or 'geostrophic' adjustment - relaxation of the atmosphere to a balanced state after the distortion of a jet stream by a weather system.

IGWs are indirectly important for synoptic scale meteorology. The atmosphere is on the synoptic scale approximately in hydrostatic equilibrium and striving towards geostrophic balance. It can be shown that the deviations from this balance result in propagating IGWs. This process is called *geostrophic adjustment* [50].

- *Rossby waves.* They are also called planetary waves and dominate the large scale weather patterns. These waves are moving westward relative to the eastward background flow so that, depending on the situation (e.g., the wavelength of the Rossby wave, strength background flow,...), they can propagate east or westwards. Planetary waves are known to be very important for the meridional transport of heat and momentum [133]. Their underlying physics principle is the conservation of potential vorticity and they are only found if a meridional variation of the Coriolis parameter is included. Often this is taken into account by the so-called β-plane analysis with $\beta = \frac{\partial f}{\partial y}(y_0)$ and y the meridional coordinate so that $f(y) = f(y_0) + \beta(y - y_0)$ [1]. Orography is known to excite Rossby waves that are stationary as the forcing is not varying in time, as explained in [50].

- *sound waves.* Sound waves have high frequencies and large phase velocities. They are characterized by compression and expansion of air parcels. Irrelevant for meteorology, they may be eliminated by filtering approximations e.g., the anelastic approximation or the hydrostatic approximation (except for the horizontal Lamb wave).

[1] This β factor follows from the derivation of the vorticity equation on a rotating sphere.

One could summarize the previous overview as follows. Rossby waves are slowly propagating planetary waves of crucial importance for weather and climate. The exact role, impact, sources,... of IGWs is still uncertain. They are relevant for geostrophic adjustment at the synoptic scales and are thought to have a more direct importance at mesoscales (e.g., clear air turbulence). The very fast propagating sound waves are not important for meteorology. This means that weather models must represent Rossby waves accurately, should preferably have an appropriate representation of IGWs (especially mesoscale models), and do not have to bother about sound waves (as long as they do not cause numerical problems in your model).

A wave can be seen as an oscillation around a balanced state and is mathematically described as

$$\hat{f}\,(k,l,m)\,e^{i(kx+ly+mz-\omega t)}, \tag{3.1}$$

with k, l, and m the wavenumber in the x-, y-, and z-direction, ω the frequency, and $\hat{f}$ the amplitude. Wavenumbers are related to the wavelength such that, for example, $k = \frac{2\pi}{\lambda_x}$. The frequency ω is a complex number whose real part $\Re(\omega)$ and imaginary part $\Im(\omega)$ determine the phase and the amplitude of the wave, respectively. The *phase velocity* $v_{ph} = \frac{\Re(\omega)}{\sqrt{k^2+l^2+m^2}}$ is the velocity the wave fronts are propagating at. The *group velocity* $\mathbf{v}_{gr} = \frac{\partial\Re(\omega)}{\partial k}\mathbf{1_x} + \frac{\partial\Re(\omega)}{\partial l}\mathbf{1_y} + \frac{\partial\Re(\omega)}{\partial m}\mathbf{1_z}$ represents the velocity at which energy is propagating. A wave is called dispersive if its phase velocity depends on the wavenumber. A function consisting of dispersive waves is characterized by a changing shape during propagation. The relation between frequency and wavenumber is called the dispersion relation. As an example, Figure 1.4 visualized the dispersion relation of the Rossby waves, sound waves, and IGWs for the equations with and without hydrostatic approximation.
It is well-known that a periodic function $f\,(x,y,z,t)$ can be decomposed into a sum of waves:

$$f\,(x,y,z,t) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\hat{f}\,(k,l,m)\,e^{i(kx+ly+mz-\omega t)}. \tag{3.2}$$

The fact that functions can be seen as an infinite sum of waves explains why studying the wave behavior is essential. The dependency of the amplitudes $\hat{f}$ on the wavenumbers results in the spectrum of the field f. Decomposition (3.2) forms the foundation for the spectral spatial discretization approach albeit one then needs to truncate the series at certain wavenumbers K, L, and M due to the limited computational capacities:

$$f\,(x,y,z,t) \approx \sum_{k=0}^{K}\sum_{l=0}^{L}\sum_{m=0}^{M}\hat{f}\,(k,l,m)\,e^{i(kx+ly+mz-\omega t)}. \tag{3.3}$$

The more wave components one adds, the closer the series approximates the function f.
Dispersion relations can be found by inserting Eq. (3.2) for each variable into the equation set describing the wave of interest. Remark that for a wave all fields (pressure, velocity,...) oscillate with the same frequency ω and have the same wavenumbers k, l, and m. Some arithmetic then results in a relation between the frequency and the wavenumber. Depending on the equation set used, different dispersion relations are found. The number of prognostic equations determines the number of waves permitted by the equation set. Transforming a prognostic equation into a diagnostic equation corresponds to filtering one wave type. As an example, consider the hydrostatic equation, which has a diagnostic vertical momentum equation (1.13). This equation set therefore filters out the sound waves, as illustrated in Figure 1.4.

3.1.2 The shallow water equations, a toy system to study the behavior of atmospheric waves

As explained in [50], the SWE can be found by introducing some extra assumptions in the hydrostatic primitive equations (Eq. (1.10), Eq. (1.11), Eq. (1.3), and Eq. (1.13)). We assume an incompressible atmosphere with density ρ. The height of the atmospheric layer $h(x,y,t)$ will change and via the hydrostatic relation this gives for the surface pressure $p(x,y,t) = \rho g h(x,y,t)$. Assuming now a 2D planar geometry instead of the spherical geometry, Eq. (1.10) and Eq. (1.11) are reduced to

$$\frac{du}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + fv = -g\frac{\partial h}{\partial x} + fv = -\frac{\partial \phi}{\partial x} + fv \text{ and} \tag{3.4}$$

$$\frac{dv}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial y} - fu = -g\frac{\partial h}{\partial y} - fu = -\frac{\partial \phi}{\partial y} - fu. \tag{3.5}$$

The formulations in height h, surface pressure p, and geopotential $\phi = gh$ are equivalent if we assume that g is a constant. Remark that the pressure gradient term does not depend on the height z. Let us assume an initial horizontal velocity field that is constant with height. This means that all horizontal velocity fields will remain independent of the height; indeed, there is not one term in the prognostic equations (3.4) and (3.5) varying along the vertical.
The incompressibility reduces the continuity equation to $\nabla \cdot \mathbf{V} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0$. From this, we can now calculate the vertical velocity at the top of the layer $w(h) = -h\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)$. By remarking that the vertical velocity at the top of the layer $w(h)$ is equal to the change of the height of the fluid $\frac{dh}{dt}$, we derive a prognostic equation for

the atmospheric height (or surface pressure or geopotential)

$$\frac{dh}{dt} = -h\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right), \tag{3.6}$$

$$\frac{dp}{dt} = -p\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right), \text{ or} \tag{3.7}$$

$$\frac{d\phi}{dt} = -\phi\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right). \tag{3.8}$$

Despite the fact that SWE are obviously too simple to be used for atmospheric modeling, they are well studied in meteorology. This can be understood as they permit a simple analysis contrary to the Euler equations and their wave solutions still show a similar behavior as the wave solutions of the more complete sets of equations. Arakawa indeed explained in [4] that the dispersion relations found for the atmospheric IGWs can be cast in the one found for the IGWs of the SWE.

Let us now derive the wave solutions of the SWE. A linear system of equations is needed in order to derive dispersion relations. We therefore first linearize the equations around a reference state $u = u'$, $v = v'$, and $h = H + h'$, where the primed variables represent small amplitude oscillations around the reference state e.g. $h' << H$. The linearized SWE in the height formulation then write as

$$\frac{\partial u'}{\partial t} = -g\frac{\partial h'}{\partial x} + fv', \tag{3.9}$$

$$\frac{\partial v'}{\partial t} = -g\frac{\partial h'}{\partial y} - fu', \text{ and} \tag{3.10}$$

$$\frac{\partial h'}{\partial t} = -H\left(\frac{\partial u'}{\partial x}+\frac{\partial v'}{\partial y}\right). \tag{3.11}$$

We have chosen a uniform reference state H for the height. However, it could be advantageous to have a reference height $H(x,y)$ depending on the horizontal position. We will come back on this in 4.5.2.
We now assume periodic oscillations, the building blocks of the Fourier decomposition (3.2),

$$\begin{aligned} u'(x,y,t) &= \hat{u}(k,l)\, e^{i(kx+ly-\omega t)}, \\ v'(x,y,t) &= \hat{v}(k,l)\, e^{i(kx+ly-\omega t)}, \text{ and} \\ h'(x,y,t) &= \hat{h}(k,l)\, e^{i(kx+ly-\omega t)}. \end{aligned}$$

As there is no z-dependence for the variables the wavevector will only have horizontal components. Inserting the previous wave forms in the linearized SWE results into the

following equation set

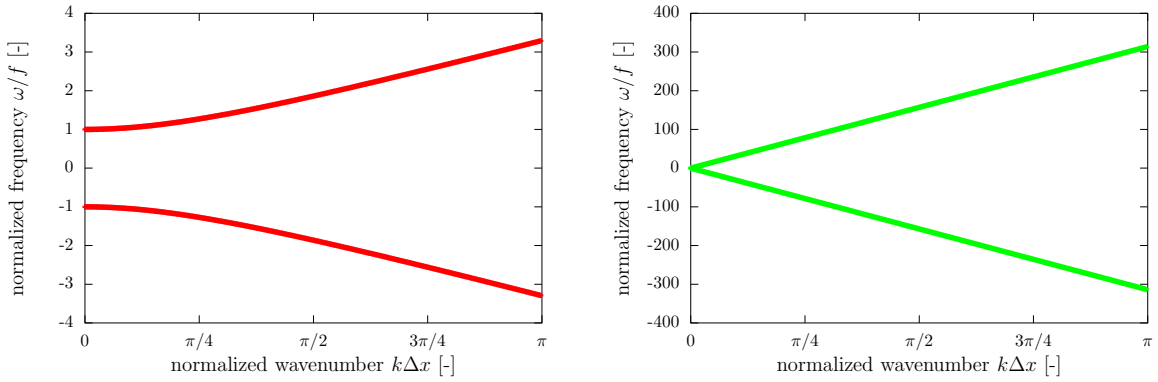
$$\begin{aligned} i\omega\hat{u} + f\hat{v} - igk\hat{h} &= 0, \\ i\omega\hat{v} - f\hat{u} - igl\hat{h} &= 0, \text{ and} \\ i\omega\hat{h} - H\left(ik\hat{u} + il\hat{v}\right) &= 0. \end{aligned}$$

Solving the previous results into the following dispersion relation for the waves described by the SWE

$$\omega_{IG}^{\pm} = \pm\sqrt{f^2 + gH\left(k^2 + l^2\right)} \quad \text{and} \quad \omega_{Ro} = 0, \tag{3.12}$$

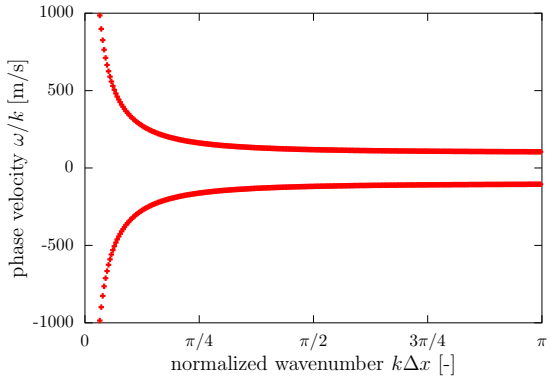
where obtaining three solutions agrees with the fact that we started from three prognostic equations. Remark that one could cast the expression (1.14) for the IGWs with the hydrostatic approximation in this SWE dispersion relation by assuming an equivalent scale height $H\left(m\right) = \frac{N^2}{g\left(m^2 + 1/\left(4H_0^2\right)\right)}$ [112].
The constants in Eq. (3.12) can be combined into the *Rossby radius of deformation* $R = \frac{\sqrt{gH}}{f}$. Arakawa [4] made an estimate for the range of Rossby radii R characteristic for the atmospheric waves relevant to NWP. In what follows, we will assume $R = 1000\,\text{km}$. The rate $\frac{R}{\Delta x}$ with Δx the grid distance will strongly influence the shape of the dispersion relation and the expressions of the phase and group velocity, as shown in Figure 3.1 for the 1D situation. Most of the studies about dispersion relations were published some decades ago when the models were ran at low resolutions, which explains why they used low rates. With the current high-resolution models it seems better to study the dispersion behavior at larger rates. Figure 3.1 compares the dispersion properties of an atmosphere with $R = 1000\,\text{km}$ for two different resolutions: $\Delta x = 1000\,\text{km}$ and $\Delta x = 10\,\text{km}$. Dispersion plots are typically presented in function of normalized wavenumber $k\Delta x$. On a mesh with grid distance Δx the shortest wave that can still be presented has a wavelength $\lambda = 2\Delta x$ and hence wavenumber $k = \frac{\pi}{\Delta x}$. One therefore has $0 < k\Delta x < \pi$ on the x-axis of Figure 3.1. Depending on the rate $\frac{R}{\Delta x}$, one ends up with plots that look completely different. All waves present in the plot for $\Delta x = 1000\,\text{km}$ are also present in the long wave limit of the plot for $\Delta x = 10\,\text{km}$. For the high-resolution case the frequency scales almost linearly with the wavenumber and one could approximate Eq. (3.12) by $\omega_{IG}^{\pm} \approx \pm\sqrt{gH\left(k^2 + l^2\right)}$.

The eigenvectors corresponding to the solutions of Eq. (3.12) have the following

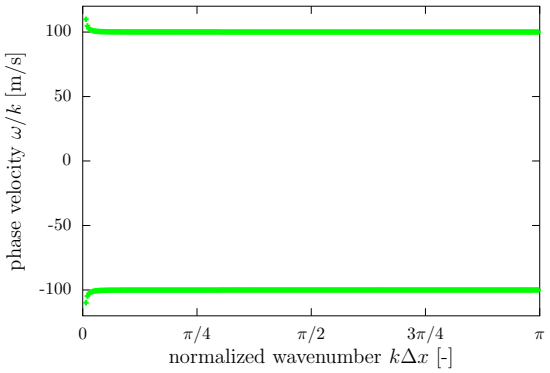


(a)

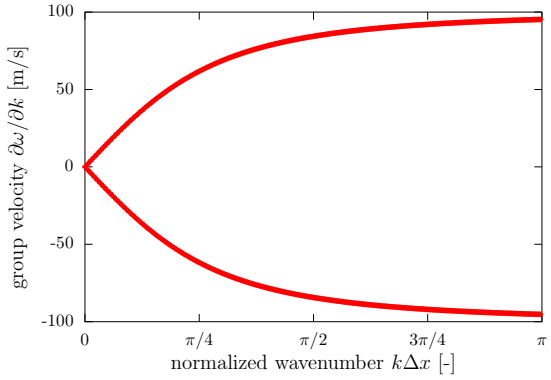
(b)



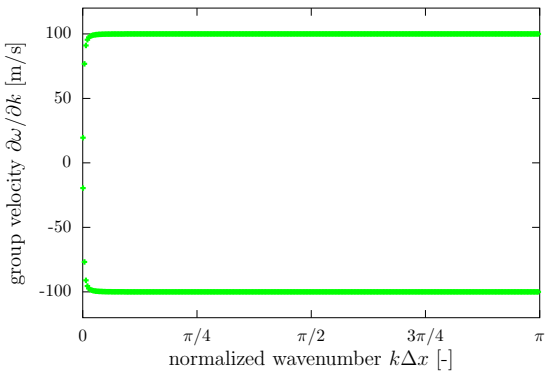
(c)



(d)



(e)



(f)

Figure 3.1: Plot of the normalized frequency (a, b), the phase velocity (c, d), and the group velocity (e, f) of the IGWs for $R = 1000\,\text{km}$. The left plots (red) show the results if $\Delta x = 1000\,\text{km}$, the right plots (green) for $\Delta x = 10\,\text{km}$. The wave propagating along the x-axis is shown ($l = 0$). On the x-axis the dimensionless wavenumber $k\Delta x$ is shown.

form

$$\mathbf{V_{ig}^{\pm}} = \begin{bmatrix} \frac{ikf - l\omega_{ig}^{\pm}}{-H(k^2+l^2)} \\ \frac{ifl + k\omega_{ig}^{\pm}}{H(k^2+l^2)} \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{V_{ro}} = \begin{bmatrix} -\frac{igl}{f} \\ \frac{igk}{f} \\ 1 \end{bmatrix}. \tag{3.13}$$

The first eigenvector corresponds to the IGWs and the second one to the stationary Rossby mode. We clearly see that the eigenvector $\mathbf{V_{ro}}$ expresses a state of geostrophic balance.

Finally, Eq. (3.4), Eq. (3.5), and Eq. (3.6) can be extended to the situation with a bottom topography h_{oro}. We then get

$$\frac{du}{dt} = -g\frac{\partial h}{\partial x} + fv, \tag{3.14}$$

$$\frac{dv}{dt} = -g\frac{\partial h}{\partial y} - fu, \text{ and} \tag{3.15}$$

$$\frac{d\left(h - h_{oro}\right)}{dt} = -\left(h - h_{oro}\right)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right). \tag{3.16}$$

In the following sections the linearized SWE will be used to study the impact of discretization on the properties of IGWs.

3.2 Impact of time discretization on IGW dispersion

The scope of this thesis is limited to SI time discretization schemes. We, therefore, only study the impact of the SI time discretization on the dispersion properties. To do so, we start from the linearized SWE (3.9), (3.10), and (3.11). Applying a 2TL SI discretization then yields

$$\frac{u'^{+} - u'^{0}}{\Delta t} = -\frac{g}{2}\left[\left(\frac{\partial h'}{\partial x}\right)^{+} + \left(\frac{\partial h'}{\partial x}\right)^{0}\right] + \frac{f}{2}\left(v'^{+} + v'^{0}\right), \tag{3.17}$$

$$\frac{v'^{+} - v'^{0}}{\Delta t} = -\frac{g}{2}\left[\left(\frac{\partial h'}{\partial y}\right)^{+} + \left(\frac{\partial h'}{\partial y}\right)^{0}\right] - \frac{f}{2}\left(u'^{+} + u'^{0}\right), \text{ and} \tag{3.18}$$

$$\frac{h'^{+} - h'^{0}}{\Delta t} = -\frac{H}{2}\left[\left(\frac{\partial u'}{\partial x}\right)^{+} + \left(\frac{\partial u'}{\partial x}\right)^{0} + \left(\frac{\partial v'}{\partial y}\right)^{+} + \left(\frac{\partial v'}{\partial y}\right)^{0}\right], \tag{3.19}$$

with ^{0} and $^+$ representing an evaluation at time t^0 and $t^+ = t^0 + \Delta t$. For the sake of simplicity, all forcing terms are treated implicitly although in reality this is often not

the case for the Coriolis force (e.g., [115]). If we now fill in the wave forms (3.1) and assume exact spatial derivatives, we obtain:

$$2i\sin\left(\frac{\omega\Delta t}{2}\right)\hat{u} - f\Delta t\cos\left(\frac{\omega\Delta t}{2}\right)\hat{v} + igk\Delta t\cos\left(\frac{\omega\Delta t}{2}\right)\hat{h} = 0, \tag{3.20}$$

$$f\Delta t\cos\left(\frac{\omega\Delta t}{2}\right)\hat{u} + 2i\sin\left(\frac{\omega\Delta t}{2}\right)\hat{v} + igl\Delta t\cos\left(\frac{\omega\Delta t}{2}\right)\hat{h} = 0, \text{ and} \tag{3.21}$$

$$ikH\Delta t\cos\left(\frac{\omega\Delta t}{2}\right)\hat{u} + ilH\Delta t\cos\left(\frac{\omega\Delta t}{2}\right)\hat{v} + 2i\sin\left(\frac{\omega\Delta t}{2}\right)\hat{h} = 0. \tag{3.22}$$

This yields the following dispersion relations

$$\omega_{IG}^{\pm} = \frac{2}{\Delta t}\arctan\left[\pm\sqrt{\frac{f^2(\Delta t)^2 + gH(\Delta t)^2(k^2+l^2)}{4}}\right] \quad \text{and} \quad \omega_{Ro} = 0. \tag{3.23}$$

Figure 3.2 visualizes the IGW dispersion relation for different time steps. We clearly notice the phase speed retardation accompanying a SI discretization, as mentioned in chapter 1. In the limit of short time steps, the SI dispersion properties converge towards the analytical SWE ones. But for long time steps, a natural choice for SI schemes, there is a large impact on the properties of the IGWs. The decrease in frequency of the shortest IGWs makes that the SI time discretized scheme remains stable even for large time steps. The distortion of the shortest IGWs is considered as harmless for the forecast, because the waves in question carry little energy [108].

3.3 Impact of space discretization on IGW dispersion

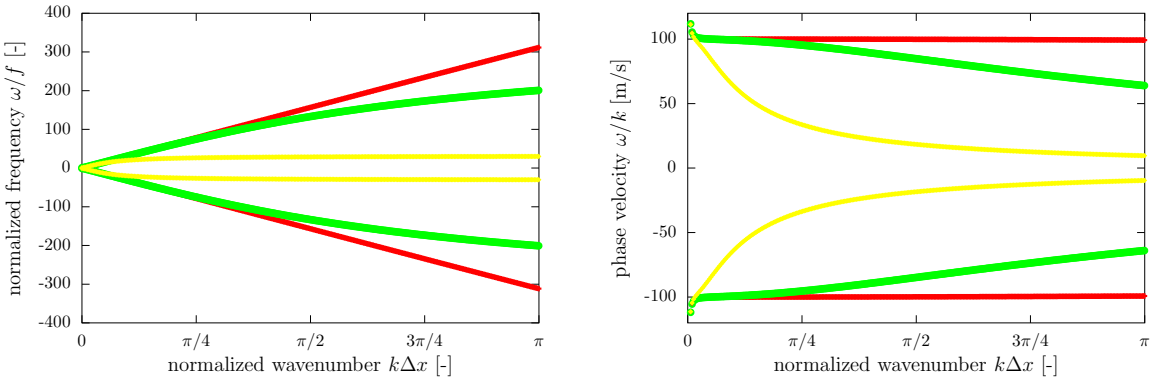
The impact of a spatial operator on a wavelike field, mathematically represented by Eq. (3.1), comes down to a multiplication of this field by a *response* function [104]. Each spatial operator will correspond to a certain response. The first-order derivative operators along the x- and y-axis correspond, for example, to the responses p_x and p_y

$$\frac{\partial u'}{\partial x} \quad \text{[cont. eq.]} \quad \leftrightarrow \quad p_x e^{i(kx+ly-\omega t)}\hat{u} \quad \text{[discr. eq. with wavesolution assumed]} \quad \text{and}$$

$$\frac{\partial u'}{\partial y} \quad \text{[cont. eq.]} \quad \leftrightarrow \quad p_y e^{i(kx+ly-\omega t)}\hat{u} \quad \text{[discr. eq. with wavesolution assumed]},$$

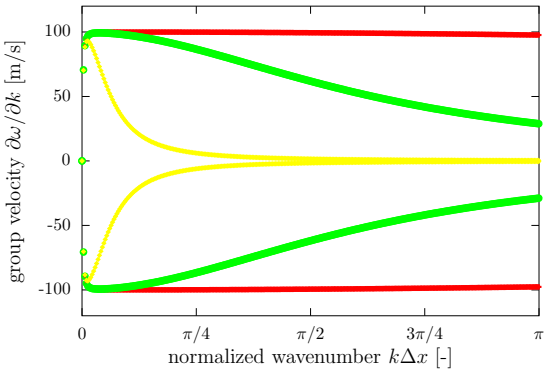
where the exact form of the response is determined by the discretization method used. Table 3.1 shows the responses corresponding to different operators for different discretization methods.

Remark that even in the absence of a spatial derivation a response p will appear, for



(a)

(b)



(c)

Figure 3.2: Plot of the normalized frequency (a), the phase velocity (b), and the group velocity (c) of the IGWs for $\lambda = 1000$ km and $\Delta x = 10$ km. A SI time discretization is used with different time steps: $\Delta t = 10$ s (red), $\Delta t = 100$ s (green), and $\Delta t = 1000$ s (yellow). The wave propagating along the x-axis is shown ($l = 0$).

Table 3.1: Responses of the second-order FD, the spectral, and the linear FE spatial discretization of a 1D wave $f = \hat{f}e^{ikx}$ for different operators. A grid distance Δx is assumed and the subscript of f denotes the position where f is evaluated. This table is based on Table 1 of [105].

operator	second-order FD	spectral	linear FE
$\mathcal{P}f$	f_x	f_x	$\frac{1}{6}\left[f_{x+\Delta x} + 4f_x + f_{x-\Delta x}\right]$
$\mathcal{P}_x f$	$\frac{1}{2\Delta x}\left[f_{x+\Delta x} - f_{x-\Delta x}\right]$	$\left(\frac{df}{dx}\right)_x$	$\frac{1}{2\Delta x}\left[f_{x+\Delta x} - f_{x-\Delta x}\right]$
$\mathcal{P}_{xx} f$	$\frac{1}{\Delta x^2}\left[f_{x+\Delta x} - 2f_x + f_{x-\Delta x}\right]$	$\left(\frac{d^2 f}{dx^2}\right)_x$	$\frac{1}{\Delta x^2}\left[f_{x+\Delta x} - 2f_x + f_{x-\Delta x}\right]$

response	second-order FD	spectral	linear FE
p	1	1	$\frac{1}{3}\left[2 + \cos(k\Delta x)\right]$
p_x	$\frac{1}{\Delta x} ik \sin(k\Delta x)$	ik	$\frac{1}{\Delta x} ik \sin(k\Delta x)$
p_{xx}	$\frac{2}{\Delta x^2}\left[\cos(k\Delta x) - 1\right]$	$-k^2$	$\frac{2}{\Delta x^2}\left[\cos(k\Delta x) - 1\right]$

example:

$$f = \frac{\partial u'}{\partial x} \quad \text{[cont. eq.]} \quad \leftrightarrow \quad p\hat{f} = p_x \hat{u} \quad \text{[discr. eq. with wavesolution assumed]}\,.$$

The function f does not undergo any derivation, but it still needs a response p. For the majority of discretization schemes (e.g., spectral, FD,...) $p = 1$. However, for Galerkin schemes with nonorthogonal basis functions this response will be nontrivial. The linear FE scheme described in appendix A illustrates this.
We will use the previously introduced responses to study the impact of space discretization on the IGWs dispersion properties. To isolate the spatial discretization impact, exact time derivation will be assumed.

3.3.1 A-grid discretization

The term A-grid was introduced by Winninghoff [136] and Arakawa [3] and is used to denote a grid where the equations formulated in terms of wind are solved on a collocation grid (left approach in Figure 3.3). Filling in the wave definition in Eq. (3.9), Eq. (3.10), and Eq. (3.11) and assuming exact time derivation, results in

$$\begin{aligned} pi\omega\hat{u} + fp\hat{v} - gp_x\hat{h} &= 0, && (3.24)\\ pi\omega\hat{v} - fp\hat{u} - gp_y\hat{h} &= 0, \text{ and} && (3.25)\\ pi\omega\hat{h} - H\left(p_x\hat{u} + p_y\hat{v}\right) &= 0. && (3.26) \end{aligned}$$

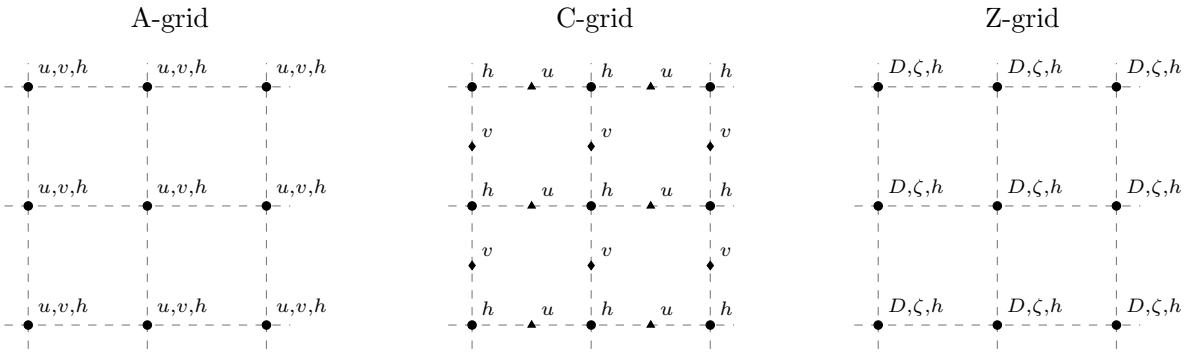


Figure 3.3: Three different grid approaches that are used in NWP. From left to right: A-grid, C-grid, and Z-grid. The figure is created with a script of Daan Degrauwe.

Solving this for the frequency ω yields the following dispersion relations:

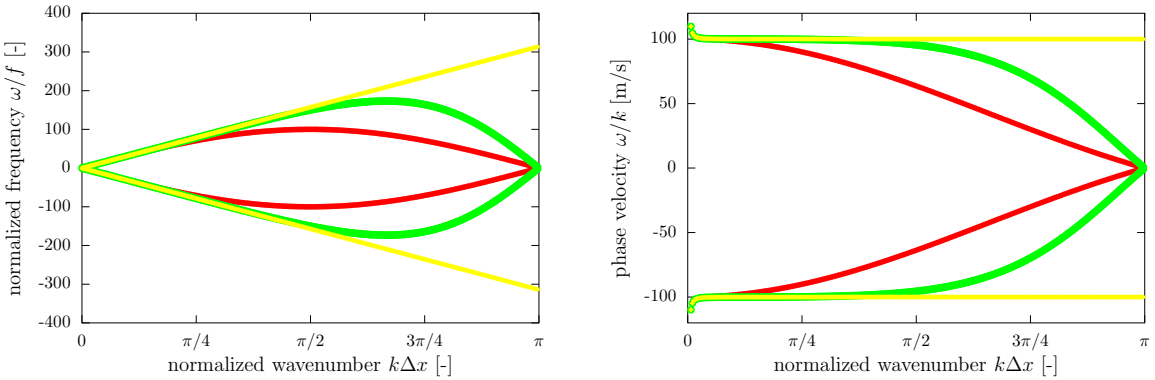
$$\omega_{IG}^{\pm} = \pm\sqrt{f^2 - gH\left(\frac{p_x^2}{p^2} + \frac{p_y^2}{p^2}\right)} \quad \text{and} \quad \omega_{Ro} = 0. \tag{3.27}$$

Remark that filling in the spectral responses, listed in Table 3.1, results in the spectral (and thus analytical) relation Eq. (3.12). The calculation of the derivative of a resolved wave is, indeed, exact if the spectral method is used. This high-order accuracy makes the spectral method very attractive.

Figure 3.4 shows the frequency, phase velocity, and group velocity for the second-order FD, linear FE, and the spectral discretization on the A-grid. The frequency ω_{IG}^{+} reaches a maximum for a certain wavenumber if one of the first two methods is applied. This should by definition correspond to a sign change of the group velocity or, put differently, from a certain wavenumber on the group velocity is reversed for the FD and FE A-grid schemes, as can be seen in Figure 3.4 (c). This means that for FD and FE schemes defined on the A-grid the shortest scale waves propagate energy opposite to the analytical direction.

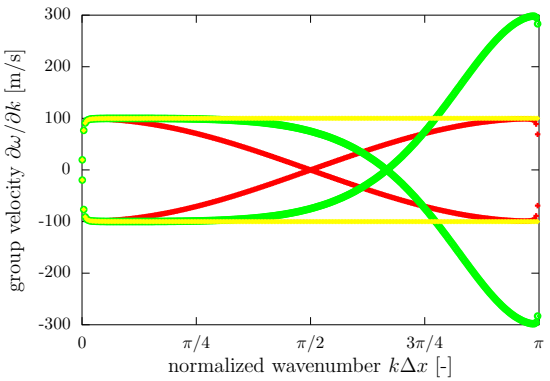
This inappropriate behavior is not related to the order of accuracy, as concluded from high-order FD [82] and spectral element [70] studies. Figure 3.5 (a) shows the dispersion of multiple FD schemes up to 8th-order accuracy. Increasing the accuracy does shift the wavenumber at which the maximal frequency is reached, but it does not avoid the inappropriate behavior of the group velocity for the shortest scale waves. This does not come as a surprise as the order of accuracy n is by definition important for the behavior in the long wave limit $k\Delta x \to 0$:

$$\epsilon = \omega_n - \omega_{sp} \approx (\Delta x)^n \longrightarrow \log\epsilon \approx n\log(\Delta x), \tag{3.28}$$



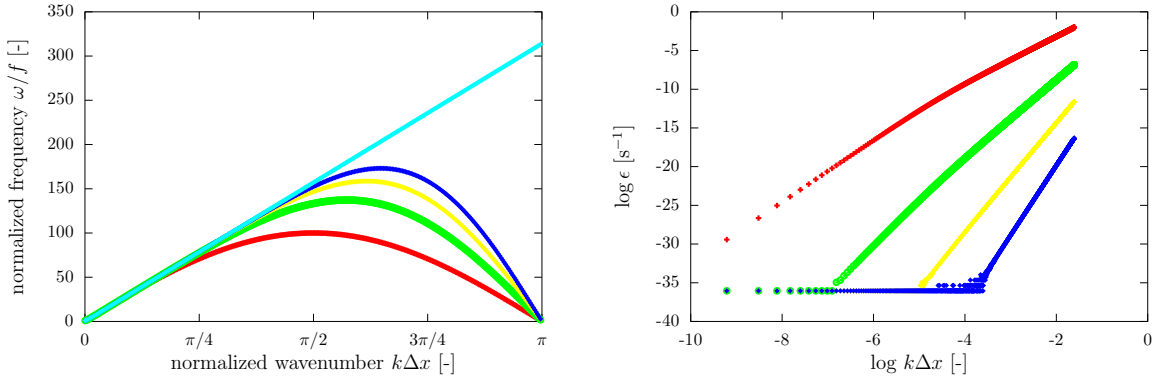
(a)

(b)



(c)

Figure 3.4: Plot of the normalized frequency (a), the phase velocity (b), and the group velocity (c) of the IGWs for $\lambda = 1000$ km and $\Delta x = 10$ km. Exact time derivation is assumed. The results are shown for different spatial discretization schemes: second-order FD (red), linear FE (green), and spectral discretization (yellow).



(a) (b)

Figure 3.5: Dispersion relation for second-order (red), fourth-order (green), sixth-order (yellow) and eight-order (blue) accurate FD in (a), where the spectral curve (cyan) is given for reference. Impact of the order of accuracy on the error $\epsilon = \omega_n - \omega_{sp}$ for the longest waves ($0 < k\Delta x < 0.002$) in (b).

with ω_n the frequency if a n^{th}-order FD scheme is used and ω_{sp} the analytical solution. To illustrate this, Figure 3.5 (b) shows a log-log plot where the slope of the logarithm of the error in the frequency $\log \epsilon = \log\left(\omega_n - \omega_{sp}\right)$ gives the order of accuracy.

The issue of the reversed group velocity of the IGWs for A-grid local discretization schemes has been known for a long time. Mesinger and Arakawa were among the first to notice and describe this [68]. In their 1D SWE study with second-order FD discretization they gave the following explanation for this unexpected behavior:

> There is, however, an important difference between this problem and the advection problem because we now have two dependent variables. We have assumed that they are both carried at every grid point [...] As far as the system [...] (*red, reference to the shallow water equations*) is concerned, however, the underlined variables in the figure depend only on other underlined variables. The same statement holds for the variables that are not underlined. Thus, the grid in the figure contains two elementary subgrids, with the solution on one of these subgrids being completely decoupled from the other.

This *decoupled grid* argument seems plausible and is often evoked in literature. Using a higher-order FD (Figure 3.5) or linear FE scheme does couple all grid points but the inappropriate behavior is still present. The explanation can be found in Eq. (3.27), which gives $\omega_{IG} = \pm f$ if the response function p_x approaches 0 for $k\Delta x \to \pi$. And this is a property of all FD and FE methods on the A-grid and therefore inhibits appropriate short scale IGW behavior.

Apart from using the spectral discretization two more approaches circumvent the A-grid geostrophic adjustment problem: use of a staggered grid organization or a formulation of the equations into a divergence-vorticity form. Both alternatives will be discussed now.
As explained earlier the modified equation approach falls out of the scope of this thesis. However it might be interesting to investigate the dispersion properties of this approach.

3.3.2 C-grid discretization

The traditional solution to the inappropriate wave dispersion consists of using a staggered grid, which is a grid where the variables are not all defined in the same grid points. Numerous authors (e.g., [68]) studied the merits of the different staggerings and in general the C-grid approach (middle approach of Figure 3.3) came out as the preferred option.

Let us derive the dispersion relation for the C-grid approach combined with second-order centered FD. The C-grid counterpart of Eq. (3.24), Eq. (3.25), and Eq. (3.26) looks like

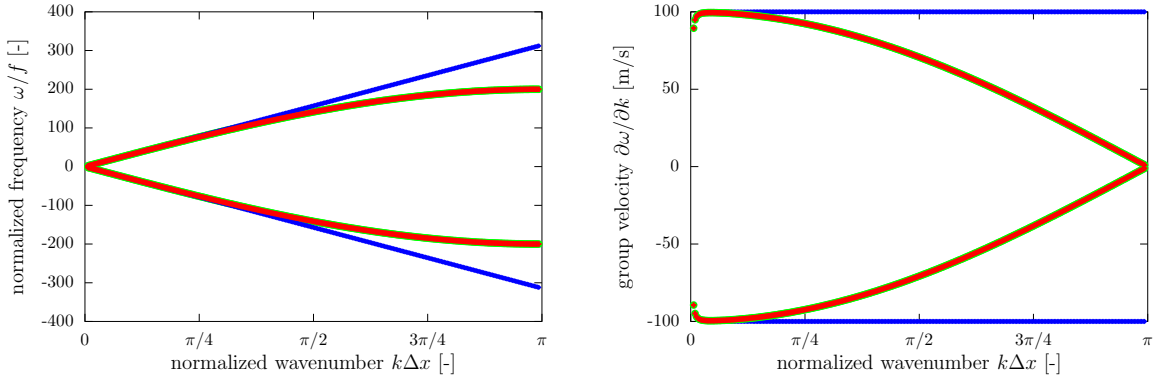
$$\begin{aligned} i\omega\hat{u} + f p_{av}\hat{v} - g p_x \hat{h} &= 0, && (3.29)\\ i\omega\hat{v} - f p_{av}\hat{u} - g p_y \hat{h} &= 0, \text{ and} && (3.30)\\ i\omega\hat{h} - H\left(p_x\hat{u} + p_y\hat{v}\right) &= 0. && (3.31) \end{aligned}$$

Eq. (3.29) and Eq. (3.30) now have a p_{av} response in front of the Coriolis terms. This is explained by the unavailability of u in the v-points and vice versa. Some averaging over the u-values in the neighboring four u-points is, for example, needed to calculate the zonal wind u in a v-point and vice versa. The corresponding response p_{av} can be expressed as

$$u^v = \frac{1}{4}\sum_i u_i^u \leftrightarrow p_{av}\hat{u} = \cos\left(\frac{k\Delta x}{2}\right)\cos\left(\frac{l\Delta y}{2}\right)\hat{u}. \tag{3.32}$$

The second-order centered spatial derivatives are now calculated by subtracting values at a distance $\frac{\Delta x}{2}$ left and right of the point instead of at a distance Δx in the A-grid approach. This makes the C-grid calculation of the derivatives more accurate by a factor 4 and yields $p_x = i\sin\left(\frac{k\Delta x}{2}\right)$. The frequency relation then takes the following form

$$\omega_{IG}^{\pm} = \pm\sqrt{f^2 p_{av}^2 - gH\left(p_x^2 + p_y^2\right)} \quad \text{and} \quad \omega_{Ro} = 0. \tag{3.33}$$



(a) (b)

Figure 3.6: Normalized frequency (a) and group velocity (b) for second-order FD on a C-grid (red), on a Z-grid (green), and the exact solution (blue) for $\lambda = 1000\,\text{km}$ and $\Delta x = 10\,\text{km}$. The C-grid and Z-grid curves are almost identical.

Figure 3.6 shows a plot of the dispersion relation of the C-grid. Energy always propagates in the correct direction. For $k\Delta x = \pi$ the energy transport becomes stationary if the C-grid approach is used.

The *mixed FE* approach is considered as the FE variant of staggering. This method uses different sets of basis functions instead of different grids for each variable. If done properly, it should be possible to obtain appropriate dispersion with this method [111, 71]. However, in practice the method looks quite cumbersome and the Z-grid scheme discussed in the next section seems more suitable for FE methods.

For waves strongly influenced by the Coriolis force special care should be taken on the C-grid. If the grid resolution is larger than the Rossby radius $\frac{R}{\Delta x} < 1$, the geostrophic adjustment is no longer appropriate [68, 86]. However, on the current high-resolution grids this does not pose a large threat. However, the discretization of the Coriolis force also impacts the Rossby waves that are meteorologically very relevant. Thuburn wrote a series of papers on how to evaluate the Coriolis terms on a C-grid in order to obtain good Rossby wave propagation [120, 122]. C-grid discretization is commonly used in NWP modeling [140, 28].

3.3.3 Z-grid discretization

One can rewrite Eq. (3.4), Eq. (3.5), and Eq. (3.6) into a formulation based on divergence D and vorticity ζ

$$\begin{aligned}
\frac{dD}{dt} &= -u\left(\frac{\partial D}{\partial x}+\frac{\partial \zeta}{\partial y}\right)-v\left(\frac{\partial D}{\partial y}-\frac{\partial \zeta}{\partial x}\right) \\
&\quad -\nabla^2\left(gh+\frac{u^2+v^2}{2}\right)+\zeta(f+\zeta), && (3.34)\\
\frac{d\zeta}{dt} &= -(\zeta+f)\,D, \text{ and} && (3.35)\\
\frac{dh}{dt} &= -HD, && (3.36)
\end{aligned}$$

with

$$\begin{aligned}
D &= \frac{\partial u}{\partial x}+\frac{\partial v}{\partial y} \text{ and} && (3.37)\\
\zeta &= \frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}. && (3.38)
\end{aligned}$$

Solving equations (3.34), (3.35), and (3.36) on a collocation grid is called the *Z-grid* approach (right approach of Figure 3.3). The vorticity and continuity equation are simple, but the divergence equation contains multiple nonlinear forcing terms, which complicates the discretization. To derive the dispersion properties, we linearize Eq. (3.34), Eq. (3.35), and Eq. (3.36) around a reference state in rest. With $u = u'$, $v = v'$, $D = D'$, $\zeta = \zeta'$, and $h = h' + H$, we find:

$$\begin{aligned}
\frac{\partial D'}{\partial t} &= -g\nabla^2 h'+f\zeta, && (3.39)\\
\frac{\partial \zeta'}{\partial t} &= -fD', \text{ and} && (3.40)\\
\frac{\partial h'}{\partial t} &= -HD'. && (3.41)
\end{aligned}$$

Remark that the linearized equation set could also have been found starting from Eq. (3.9), Eq. (3.10), and Eq. (3.11) and assuming

$$\begin{aligned}
D' &= \frac{\partial u'}{\partial x}+\frac{\partial v'}{\partial y} \text{ and} && (3.42)\\
\zeta' &= \frac{\partial v'}{\partial x}-\frac{\partial u'}{\partial y}. && (3.43)
\end{aligned}$$

Assuming wavelike solutions for Eq. (3.39), Eq. (3.40), and Eq. (3.41) then results in the following dispersion relation

$$\omega_{IG}^{\pm} = \pm\sqrt{f^2 - gH\left(\frac{p_{xx}}{p^2} + \frac{p_{yy}}{p^2}\right)} \quad \text{and} \quad \omega_{Ro} = 0. \tag{3.44}$$

The dispersion relation obtained with the Z-grid approach is represented by the green curve in Figure 3.6. Reminding $\sin^2(k\Delta x) = 2[\cos(k\Delta x) - 1]$, one understands why the second-order FD dispersion relation is nearly identical to the Z-grid one. The difference is limited to the Coriolis term (no averaging is needed for the Z-grid).
The terms responsible for the geostrophic adjustment differ between the A-grid and Z-grid scheme. If one goes from the A-grid system to Z-grid, $\frac{\partial h}{\partial x}$ and $\frac{\partial h}{\partial y}$ in the momentum equation and $\nabla \cdot \mathbf{v}$ in the continuity equation are replaced by $\nabla^2 h$ in the divergence equation and divergence D in the continuity equation. Removing the first-order derivatives by reformulating the equations in terms of divergence and vorticity is the reason why a Z-grid discretization results in a more accurate representation of geostrophic adjustment.
The term Z-grid was introduced by Randall [86], but the merits of the equations formulated in divergence-vorticity form were already known long before [104, 134]. Rossby wave propagation does not pose any problem to the Z-grid methodology, as concluded by a β-plane study of Neta [76].
Many schemes need the wind velocities at every time step, for example, SL schemes in order to calculate the departure points. For a Z-grid scheme the wind velocities are no longer prognostic variables, but they can be retrieved by solving a diagnostic equation every time step. This can be done using an inverted version of (3.37) and (3.38), which will be denoted as the DZ wind reconstruction method. Or one could solve a Poisson problem in order to calculate the wind velocities. There are two different ways to construct the Poisson equation. One could directly relate the wind velocities to divergence and vorticity (PO1 reconstruction)

$$\nabla^2 u = \frac{\partial D}{\partial x} - \frac{\partial \zeta}{\partial y} \text{ and} \tag{3.45}$$

$$\nabla^2 v = \frac{\partial \zeta}{\partial x} + \frac{\partial D}{\partial y} \tag{3.46}$$

or one could use the intermediate potential velocity χ and streamfunction ψ (PO2 reconstruction)

$$\nabla^2 \chi = D \text{ and} \tag{3.47}$$

$$\nabla^2 \psi = \zeta. \tag{3.48}$$

The horizontal velocity components u and v are then restored using the Helmholtz theorem:

$$u = \frac{\partial \chi}{\partial x} - \frac{\partial \psi}{\partial y} \text{ and} \tag{3.49}$$

$$v = \frac{\partial \psi}{\partial x} + \frac{\partial \chi}{\partial y}. \tag{3.50}$$

Solving this diagnostic problem on all levels at every time step is the price to pay for the appropriate dispersion of the Z-grid. The Z-grid method will be discussed in more detail in the next chapter.

3.4 Conclusion

This chapter reviewed how waves that are solutions of our analytical equations are influenced by the numerical methods used. We studied this problem by using the SWE, a simplified set of equations that still encompasses most of the properties of the real waves. It was shown that the spectral method gives always rise to the analytical dispersion relations of the IGWs. However, local discretization methods that solve the equations formulated in terms of wind velocity on a collocation grid, called the A-grid approach, result in inappropriate geostrophic adjustment. The short scale end of the IGW spectrum will propagate energy in the wrong direction. This can be avoided by using a staggered grid formulation (typically C-grid) or by reformulating the equations in terms of vorticity and divergence on an unstaggered grid, known as the Z-grid approach. Following the modularity constraint, one should keep the collocation grid approach of the current ALADIN model. The Z-grid methodology, therefore, seems a good approach to introduce a local spatial discretization method. In the following chapter we will study such an implementation in detail.

It is interesting to note that there is no consensus about the importance of the A-grid dispersion. One may argue that the shortest scales, which is that part of the spectrum where the problematic IGW behavior is situated, are dominated by physics parameterizations and not by dynamics. Moreover, most models use diffusion or truncation at the shortest scales and if not, one could control any short scale noise by adding some diffusion at the cost of losing some accuracy [134, 82, 70]. A study on the importance of IGW dispersion within a full NWP model would be valuable (in chapter 5 we give it a try) but in the meantime it seems more safe to take care about appropriate dispersion.

SWE study of local horizontal spatial discretization Z-grid schemes

4

Chapter 3 shows that appropriate IGW dispersion can be obtained by three different approaches:

1. schemes based on a spectral horizontal spatial discretization,
2. schemes where the variables are defined on a staggered grid, such as the C-grid, and
3. Z-grid schemes that solve the equations formulated in divergence and vorticity on a collocation grid.

The ALADIN model is currently based on a spectral SISL dynamical core. If one would like to test an alternative local discretization approach without sacrificing the appropriate IGW dispersion, one should apply option 2 or 3. However, the Z-grid approach is preferred as it permits to keep a formulation on an unstaggered grid, which is one of the constraints discussed in chapter 2. Moreover, option 2 combined with a SL treatment of advection would treble the number of trajectory calculations. This chapter will, therefore, focus on the formulation and properties of SISL Z-grid schemes.
In this chapter the SWE are used to illustrate how a SISL Z-grid scheme works. The choice for the SWE keeps the numerics tractable while resulting in IGW dispersion relations similar to the ones obtained with the hydrostatic set of equations. To permit a comparison with the current algorithmics, section 4.1 applies the 2TL SISL ALADIN discretization on the 2D SWE. Thereafter, SISL Z-grid schemes combined with local spatial discretization methods that were proposed in literature are investigated. Despite the excellent dispersion relations found in the previous chapter, some issues pop-up while applying the Z-grid method. First, symmetry of the spatial discretization between the explicit and implicit terms will turn out to be crucial. Section 4.2 presents a dispersion analysis and section 4.3 proposes a solution, which is tested numerically. However, a second problem, which is related to the eigenvectors, is diagnosed in section 4.4. An eigenvalue study of a numerical scheme turns out to be insufficient to guarantee an appropriate representation of atmospheric waves. Section 4.5 extends

Parts of this chapter are published as [15].

the conclusions from the previous sections to the nonlinear case, by including advection and orography. The opportunity offered by local methods to include orography terms in the implicit part of the solver is investigated. Finally, some conclusions are presented together with a discussion about the findings in this chapter.

4.1 2TL SISL ALADIN approach applied on 2D SWE

The time step organization of the ALADIN model was discussed in 2.3. Let us concretize some of these steps by applying its discretization approach to the 2D SWE with orography Eq. (3.14) - (3.16). First, we need to choose a reference state and linearize the equations around this reference atmosphere. Let us define the reference state as follows: $u_{ref} = 0$, $v_{ref} = 0$, and $h_{ref} = H$. Linearization, a SL treatment of advection, and an implicit 2TL discretization of the linear terms then yield

$$\frac{u_A^+ - u_D^0}{\Delta t} = -\frac{g}{2}\left[\left(\frac{\partial h}{\partial x}\right)_A^+ + \left(\frac{\partial h}{\partial x}\right)_D^0\right] + \frac{f}{2}\left(v_A^+ + v_D^0\right), \tag{4.1}$$

$$\frac{v_A^+ - v_D^0}{\Delta t} = -\frac{g}{2}\left[\left(\frac{\partial h}{\partial y}\right)_A^+ + \left(\frac{\partial h}{\partial y}\right)_D^0\right] - \frac{f}{2}\left(u_A^+ + u_D^0\right), \text{and} \tag{4.2}$$

$$\begin{aligned}\frac{h_A^+ - h_D^0}{\Delta t} - \frac{h_{oro,A} - h_{oro,D}}{\Delta t} &= -\frac{H}{2}\left[\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)_A^+ + \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)_D^0\right] \\ &+ \left[(h_{oro} - h)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\right]_?^? .\end{aligned} \tag{4.3}$$

Despite the wind and height in the previous equations being actually perturbations on the reference state, primes were omitted not to overload the notation. The notations $()_D^0$ and $()_A^+$ express an evaluation in the departure points at time t^0 and in the arrival points at time $t^+ = t^0 + \Delta t$. We use the $[\ ]_?^?$ notation to point out that different methods can be used to evaluate the nonlinear terms. Collecting the unknown variables in the left hand side and the explicit dynamics and nonlinear terms in the right hand side

results in

$$u_A^+ + \frac{g\Delta t}{2}\left(\frac{\partial h}{\partial x}\right)_A^+ - \frac{f\Delta t}{2}v_A^+ = R_u = u_D^0 - \frac{g\Delta t}{2}\left(\frac{\partial h}{\partial x}\right)_D^0 + \frac{f\Delta t}{2}v_D^0, \tag{4.4}$$

$$v_A^+ + \frac{g\Delta t}{2}\left(\frac{\partial h}{\partial y}\right)_A^+ + \frac{f\Delta t}{2}u_A^+ = R_v = v_D^0 - \frac{g\Delta t}{2}\left(\frac{\partial h}{\partial y}\right)_D^0 - \frac{f\Delta t}{2}u_D^0, \text{ and} \tag{4.5}$$

$$\begin{aligned} h_A^+ + \frac{H\Delta t}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)_A^+ &= R_h = h_D^0 - \frac{H\Delta t}{2}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)_D^0 + h_{oro,A} - h_{oro,D} \\ &+ \Delta t\left[(h_{oro} - h)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right)\right]_?^?. \end{aligned} \tag{4.6}$$

We recognize the structure of Eq. (2.3), although without physics parameterization tendencies. The right hand side terms R_u, R_v, and R_h are known (evaluations based on values at t^0) and calculated in grid point space (step 5 in Table 2.2). Then the LBC coupling of these right hand side terms (step 7) is done [87] before transforming the fields to spectral space (step 8).
The equations are then reformulated in terms of divergence D and vorticity ζ

$$D_A^+ + \frac{g\Delta t}{2}\left(\frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2}\right)_A^+ - \frac{f\Delta t}{2}\zeta_A^+ = R_D = \frac{\partial R_u}{\partial x} + \frac{\partial R_v}{\partial y}, \tag{4.7}$$

$$\zeta_A^+ + \frac{f\Delta t}{2}D_A^+ = R_\zeta = \frac{\partial R_v}{\partial x} - \frac{\partial R_u}{\partial y}, \text{ and} \tag{4.8}$$

$$h_A^+ + \frac{H\Delta t}{2}D_A^+ = R_h. \tag{4.9}$$

In spectral space the previous equations are written as:

$$\begin{aligned} \widehat{D}^+(k,l) - \frac{g\Delta t}{2}\left(k^2 + l^2\right)\widehat{h}^+(k,l) - \frac{f\Delta t}{2}\widehat{\zeta}^+(k,l) &= \widehat{R}_D(k,l) \\ &= ik\widehat{R}_u(k,l) + il\widehat{R}_v(k,l), \end{aligned} \tag{4.10}$$

$$\begin{aligned} \widehat{\zeta}^+(k,l) + \frac{f\Delta t}{2}\widehat{D}^+(k,l) &= \widehat{R}_\zeta(k,l) \\ &= ik\widehat{R}_v(k,l) - il\widehat{R}_u(k,l), \text{ and} \end{aligned} \tag{4.11}$$

$$\widehat{h}^+(k,l) + \frac{H\Delta t}{2}\widehat{D}^+(k,l) = \widehat{R}_h(k,l), \tag{4.12}$$

where $\widehat{f}(k,l)$ denotes the spectral coefficient corresponding to the wave with wavenumbers k and l. Finally, by some algebraic manipulations of Eq. (4.10) - (4.12) one Helm-

holtz equation for the spectral coefficients of the divergence $\widehat{D}^{+}(k,l)$ can be derived

$$\left[1+\left(\frac{f\Delta t}{2}\right)^{2}+gH\left(k^{2}+l^{2}\right)\right]\widehat{D}^{+}(k,l)=\widehat{R}_{D}(k,l)+\frac{f\Delta t}{2}\widehat{R}_{\zeta}(k,l)$$
$$+\frac{g\Delta t}{2}\left(k^{2}+l^{2}\right)\widehat{R}_{h}(k,l)\,. \tag{4.13}$$

The spectral representations of vorticity $\widehat{\zeta}^{+}(k,l)$ and height $\widehat{h}^{+}(k,l)$ are then found by substitution of the updated divergence coefficients in Eq. (4.11) and Eq. (4.12). Finally, the spectral components of the wind velocities are retrieved from divergence and vorticity.

In fact, this solution procedure makes the ALADIN approach a Z-grid scheme. One may wonder why a spectral model solves the equations formulated in terms of divergence and vorticity. This transformation cannot be explained by the quest for appropriate geostrophic adjustment because this is guaranteed by the spectral discretization. It could maybe be traced back to [13] where it was written:

> An advantage of the equations expressed as above [vorticity-divergence formulation] is that they are concise. The actual number of fields involved in nonlinear products is less than is required in the form that employs U and V as the wind prognostics.

However, the SL methodology can be combined more easily with the u-v than the D-ζ formulation. Ritchie therefore proposed to have a SISL time discretization on the u-v equations and cross differentiate them in spectral space to end up with equations for divergence and vorticity [88]. Temperton remarked in [114] that for a spectral model this approach is identical to a direct discretization of the equations formulated in D-ζ:

> This illustrates one of the nice features of the spectral method: although there is certainly more than one way to organize the computation, there is fundamentally no argument about what is to be done.

Due to the analytical character of the spectral derivations, a numerical scheme based on the spectral method is not sensitive to manipulations of the equations. This is an important advantage of the spectral method. Using local methods for the spatial discretization will annihilate this property, as illustrated in the next section.

4.2 Importance of symmetry in SI Z-grid schemes

In this section various 2TL SISL Z-grid schemes proposed before in literature are examined. We can distinguish two research groups that developed a SISL Z-grid approach with a FD or FE spatial discretization:

1. During the seventies and eighties scientists of the Canadian Atmospheric Environment Service were working on unstaggered FE schemes. They proposed a SISL Z-grid scheme even before the introduction of the Z-grid nomenclature by Randall [86]. The discretization approaches they presented in different publications were always quite similar. The 3TL SISL scheme of the Staniforth and Temperton paper of 1986 [105] is used as the prototype of their approach because in this paper a dispersion analysis was undertaken[1]. For unspecified reasons this research group gave up the Z-grid approach and switched in the early nineties to staggered local SISL schemes [21].
2. The Russian SL-AV (Semi-Lagrangian Absolute Vorticity) model, developed by Tolstykh, is a second example of a vorticity-divergence based model with SISL discretization [123, 124]. The spatial discretization used by SL-AV is based on so-called compact FD, which turns out to be very similar to linear FE. The evaluation of longitudinal derivatives is still done in Fourier space, however the responses are modified to correspond to the compact FD approach. Meridional derivatives are calculated in grid point formulation. This dual treatment is motivated by the limited computational scalability of the Legendre transforms.

The following study will focus on the schemes proposed in [105] (ST86) and [124] (TS12). First, their numerics are illustrated by applying them on the linearized SWE in a 1D Cartesian context assuming no changes along the y-direction ($\frac{\partial}{\partial y} = 0$). Although this framework is simple, it includes all the terms needed to illustrate the main features of the different schemes. Then, a dispersion analysis of these schemes is presented and, finally, some unexpected features are related to asymmetries present in the SI Z-grid schemes. At the end of this section, one should be convinced about the subtleties related to the use of local spatial discretization schemes in combination with SI Z-grid schemes.

4.2.1 Construction of 2TL SISL Z-grid schemes

The time discretization of the evolution equations (3.34)-(3.36) formulated in terms of divergence D and vorticity ζ would result in many nonlinear terms, especially for the

[1] One year later a 2TL SISL variant was published [113], but this paper did not include a dispersion study.

divergence equation. Moreover, Ritchie [88] and Robert [106] suggested care should be taken when a SL treatment is used for the equations formulated in terms of divergence and vorticity. Ritchie wrote [88]:

> Since Lagrangian derivatives arise more naturally in the momentum (U-V) form of the horizontal equations of motion than they do in the ζ-D equations, U-V formulations are more appropriate for SL models.

In [106] Staniforth told about André Robert:

> At first André thought that all that was required was to apply a SL treatment to the other two prognostic equations, but this was also inaccurate and can be attributed to divergence not being a naturally advected quantity. However, momentum is, and this led him to a reformulation using the primitive (i.e., undifferentiated) form of the equations.

For the Z-grid schemes presented in ST86 and TS12 one therefore replaced the time discretization of the divergence equation by a discrete divergence equation derived from the time discretized momentum equations. For the linearized SWE with fields constant along the y-direction this yields

$$u_A^+ + \frac{g\Delta t}{2}\left(\frac{\partial h}{\partial x}\right)_A^+ - \frac{f\Delta t}{2}v_A^+ = R_u, \tag{4.14}$$

$$v_A^+ + \frac{f\Delta t}{2}u_A^+ = R_v, \tag{4.15}$$

$$h_A^+ + \frac{H\Delta t}{2}D_A^+ = R_h, \text{ and} \tag{4.16}$$

$$\zeta_A^+ + \frac{f\Delta t}{2}D_A^+ = R_\zeta. \tag{4.17}$$

The right hand sides,

$$R_u = u_D^0 - \frac{g\Delta t}{2}\left(\frac{\partial h}{\partial x}\right)_D^0 + \frac{f\Delta t}{2}v_D^0, \tag{4.18}$$

$$R_v = v_D^0 - \frac{f\Delta t}{2}u_D^0, \tag{4.19}$$

$$R_h = h_D^0 - \frac{H\Delta t}{2}D_D^0, \text{ and} \tag{4.20}$$

$$R_\zeta = \zeta_D^0 - \frac{f\Delta t}{2}D_D^0, \tag{4.21}$$

are evaluated at time t^0 in the departure points by using a suitable interpolation formula. This approach is similar to the ALADIN organization discussed earlier. An extra SL interpolation is needed to evaluate the explicit part R_ζ of the vorticity equation. However, it is possible to circumvent this by constructing the discrete vorticity equa-

tion from the discrete momentum equations as it is done for the discrete divergence equation. The Coriolis term is treated in an implicit way. Nevertheless, there are other options, such as advecting the Coriolis term as done in the SL-AV model [123, 124].
By differentiating Eq. (4.14), one can construct the discretized divergence equation

$$D_A^+ + \beta_f \frac{g\Delta t}{2} \left(\frac{\partial^2 h}{\partial x^2} \right)_A^+ = \beta_f \frac{\partial R_u}{\partial x} + \beta_f \frac{f\Delta t}{2} \frac{\partial R_v}{\partial x}, \tag{4.22}$$

with $\beta_f = \left[1 + \left(f^2 \Delta t^2\right)/4\right]^{-1}$. Combining the discretized divergence equation (4.22) and the discretized continuity equation (4.16) then yields a Helmholtz equation for the height

$$\left[1 - \beta_f \frac{gH\left(\Delta t\right)^2}{4} \frac{\partial^2}{\partial x^2}\right] h_A^+ = -\beta_f \frac{H\Delta t}{2} \left[\frac{\partial R_u}{\partial x} + \frac{f\Delta t}{2} \frac{\partial R_v}{\partial x}\right] + R_h. \tag{4.23}$$

Back substitution of the updated height field into Eq. (4.16) and Eq. (4.17) returns the divergence and vorticity fields. In ST86 the wind velocities are then retrieved by solving a discretized Poisson equation based on the velocity potential and streamfunction given by Eq. (3.47) - (3.48) and Eq. (3.49) - (3.50). This wind reconstruction procedure was denoted as PO2 reconstruction in 3.3.3. The scheme proposed by TS12 makes use of DZ reconstruction.
One can combine many space discretization methods with this SISL Z-grid approach. The Canadian Atmospheric Environment Service used linear FE, in TS12 compact FD were implemented. To cover all methods the spatial discretization will be represented by general operators. The mass operator is denoted by $\mathcal{P}$ and will be the identity operator in case of a FD or spectral discretization. The discrete counterparts of the first-order and second-order derivative operators are represented by $\mathcal{P}_x$ and $\mathcal{P}_{xx}$, respectively. By using the previously introduced operators, spatial discretization of (4.16), (4.17), and (4.22) yields

$$\mathcal{P}h^+ + \frac{H\Delta t}{2} \mathcal{P}D^+ = \mathcal{P}R_h, \tag{4.24}$$

$$\mathcal{P}\zeta^+ + \frac{f\Delta t}{2} \mathcal{P}D^+ = \mathcal{P}R_\zeta, \text{ and} \tag{4.25}$$

$$\mathcal{P}D^+ + \beta_f \frac{g\Delta t}{2} \mathcal{P}_{xx} h^+ = \beta_f \mathcal{P}_x R_u + \beta_f \frac{f\Delta t}{2} \mathcal{P}_x R_v, \tag{4.26}$$

with

$$\mathcal{P}R_u = \mathcal{L}\left(\mathcal{P}u^0 - \frac{g\Delta t}{2}\mathcal{P}_x h^0 + \frac{f\Delta t}{2}\mathcal{P}v^0\right), \tag{4.27}$$

$$\mathcal{P}R_v = \mathcal{L}\left(\mathcal{P}v^0 - \frac{f\Delta t}{2}\mathcal{P}u^0\right), \tag{4.28}$$

$$\mathcal{P}R_h = \mathcal{L}\left(\mathcal{P}h^0 - \frac{H\Delta t}{2}\mathcal{P}D^0\right), \text{ and} \tag{4.29}$$

$$\mathcal{P}R_\zeta = \mathcal{L}\left(\mathcal{P}\zeta^0 - \frac{f\Delta t}{2}\mathcal{P}D^0\right). \tag{4.30}$$

The discrete operator $\mathcal{L}$ represents the SL interpolation to the departure points. The time stepping is organized as follows:

1. Calculate the right hand sides R_u, R_v, R_h, and R_ζ by evaluating Eq. (4.27) - (4.30).
2. Solve the Helmholtz equation for h^+ derived from Eq. (4.24) and Eq. (4.26):

$$\left[\mathcal{P} - \beta_f \frac{gH(\Delta t)^2}{4}\mathcal{P}_{xx}\right] h^+ = -\beta_f \frac{H\Delta t}{2}\mathcal{P}_x\left(R_u + \frac{f\Delta t}{2}R_v\right) + \mathcal{P}R_h. \tag{4.31}$$

   Afterwards, Eq. (4.24) and Eq. (4.25) are used to calculate D^+ and ζ^+.

3. Update the wind velocities u^+ and v^+. In literature different methods are used to reconstruct the wind from divergence and vorticity. In ST86 and [123] PO2 reconstruction leads to

$$\mathcal{P}u^+ = \mathcal{P}_x\mathcal{P}_{xx}^{-1}\mathcal{P}D^+ \text{ and} \tag{4.32}$$

$$\mathcal{P}v^+ = \mathcal{P}_x\mathcal{P}_{xx}^{-1}\mathcal{P}\zeta^+, \tag{4.33}$$

   where $\mathcal{P}_{xx}^{-1}$ represents the discrete counterpart of the continuous inverse operator of the second-order derivative operator $\mathcal{P}_{xx}$. On the other hand, in TS12 the DZ reconstruction method is applied

$$\mathcal{P}_x u^+ = \mathcal{P}D^+ \text{ and} \tag{4.34}$$

$$\mathcal{P}_x v^+ = \mathcal{P}\zeta^+. \tag{4.35}$$

   Care must be taken in case operator inversions, represented symbolically by $^{-1}$, are needed. Additional constraints, such as for example fixing an average value or assuming periodicity, are needed to ensure that $\mathcal{P}_x$ or $\mathcal{P}_{xx}$ are not singular and can be inverted.

Finally, there is the PO1 update method (not used in literature to the authors' knowledge) to retrieve the wind velocities. This approach makes use of a discretized version

of the Poisson equation (3.45) and (3.46):

$$\mathcal{P}_{xx}u^{+} = \mathcal{P}_x D^{+} \text{ and} \tag{4.36}$$

$$\mathcal{P}_{xx}v^{+} = \mathcal{P}_x \zeta^{+}. \tag{4.37}$$

4.2.2 Dispersion analysis of SI Z-grid schemes

A dispersion analysis is undertaken to estimate the ability of the previous schemes to represent geostrophic adjustment. Advection is not really important herein, hence the SL interpolation operator $\mathcal{L}$ is put equal to the identity operation. The scheme represented by Eq. (4.24) - (4.26) and Eq. (4.27) - (4.30) can be reformulated symbolically into the following matrix form

$$\begin{bmatrix} \mathcal{P} & 0 & \beta_f \frac{g\Delta t}{2}\mathcal{P}_{xx} \\ \frac{f\Delta t}{2}\mathcal{P} & \mathcal{P} & 0 \\ \frac{H\Delta t}{2}\mathcal{P} & 0 & \mathcal{P} \end{bmatrix} \begin{bmatrix} D^{+} \\ \zeta^{+} \\ h^{+} \end{bmatrix} = \begin{bmatrix} \beta_f \mathcal{P}_x & \beta_f \frac{f\Delta t}{2}\mathcal{P}_x & 0 & 0 \\ 0 & 0 & 0 & \mathcal{P} \\ 0 & 0 & \mathcal{P} & 0 \end{bmatrix}$$

$$\times \begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{g\Delta t}{2}\mathcal{P}^{-1}\mathcal{P}_x & 0 & 0 \\ -\frac{f\Delta t}{2} & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -\frac{H\Delta t}{2} & 0 \\ 0 & 0 & 0 & -\frac{f\Delta t}{2} & 1 \end{bmatrix} \begin{bmatrix} u^0 \\ v^0 \\ h^0 \\ D^0 \\ \zeta^0 \end{bmatrix}. \tag{4.38}$$

Given a relation to retrieve wind velocity from divergence and vorticity, one could rewrite the previous expression into a matrix equation with only one state vector $X = [u, v, h]$

$$\mathbf{M_i} X^{+} = \mathbf{M_e} X^{0}, \tag{4.39}$$

with

$$\mathbf{M_i} = \begin{bmatrix} \mathcal{P}\mathcal{P}_x^{*} & 0 & \beta_f \frac{g\Delta t}{2}\mathcal{P}_{xx} \\ \frac{f\Delta t}{2}\mathcal{P}\mathcal{P}_x^{*} & \mathcal{P}\mathcal{P}_x^{*} & 0 \\ \frac{H\Delta t}{2}\mathcal{P}\mathcal{P}_x^{*} & 0 & \mathcal{P} \end{bmatrix} \text{ and} \tag{4.40}$$

$$\mathbf{M_e} = \begin{bmatrix} \alpha_f \mathcal{P}_x & \beta_f f\Delta t \mathcal{P}_x & -\beta_f \frac{g\Delta t}{2}\mathcal{P}_x\mathcal{P}^{-1}\mathcal{P}_x \\ -\frac{f\Delta t}{2}\mathcal{P}\mathcal{P}_x^{*} & \mathcal{P}\mathcal{P}_x^{*} & 0 \\ -\frac{H\Delta t}{2}\mathcal{P}\mathcal{P}_x^{*} & 0 & \mathcal{P} \end{bmatrix}, \tag{4.41}$$

where, for conciseness, we have $\alpha_f = \beta_f \left[1 - \left(\Delta t^2 f^2\right)/4\right]$. The operator $\mathcal{P}_x^*$ is a first-order derivative operator that represents the used transformation from the vorticity and divergence variables to the wind velocity components: $D = \mathcal{P}_x^* u$ and $\zeta = \mathcal{P}_x^* v$. It is defined to make the distinction between the different ways to retrieve the wind velocities. For the PO2 reconstruction proposed in ST86, it yields $\mathcal{P}_x^* = \mathcal{P}^{-1}\mathcal{P}_{xx}\mathcal{P}_x^{-1}\mathcal{P}$, whereas the DZ method with $\mathcal{P}_x^* = \mathcal{P}^{-1}\mathcal{P}_x$ is used in the TS12 scheme. For completeness, $\mathcal{P}_x^*$ would be equal to $\mathcal{P}_x^{-1}\mathcal{P}_{xx}$ if the wind velocity field was updated using (4.36) and (4.37).

The amplification matrix of the scheme, that is the matrix $\mathbf{A}$ one should multiply the state vector with to progress by one time step, is derived easily from Eq. (4.39). It writes as

$$\mathbf{A} = {\mathbf{M_i}}^{-1}\mathbf{M_e}, \tag{4.42}$$

and characterizes the properties of the numerical scheme completely. The dispersion properties are investigated by assuming the fields u, v, and h to be wavelike disturbances around the reference state such that the operators in Eq. (4.42) can be replaced by their responses, as given in Table 3.1.

The three complex eigenvalues λ_A of the amplification matrix $\mathbf{A}$ (one for the Rossby wave and two for the IGWs) contain all information about the propagation of the discretized eigenmodes:

- The amplification factor after one time step Δt is given by $|\lambda_A|$.
- The discretized frequency is given by $\omega = \arctan\left\{\frac{\Im(\lambda_A)}{\Re(\lambda_A)}\right\}$ with $\Re(\lambda_A)$ the real part and $\Im(\lambda_A)$ the imaginary part of the eigenvalue.

The dispersion properties of the 1D variants of the SISL Z-grid schemes proposed in ST86[2] and TS12 are shown in Figure 4.1 for two different time step values. The only difference between the two schemes is the different wind reconstruction method, or thus the different response for $\mathcal{P}_x^*$. The plots are made by assuming linear FE and second-order FD for the horizontal spatial discretization. However, in TS12 fourth-order accurate spatial discretization methods are used. It was verified -but is not shown here- that the order of accuracy of the spatial discretization does not influence the qualitative dispersion behavior. For the sake of simplicity, we limit the discussion in this paper to second-order versions of the schemes presented in ST86 and TS12. Some conclusions can be drawn based on Figure 4.1:

- The normalized frequency of the IGWs found if a second-order FD or a linear FE discretization is used is far from the spectral and time continuous Z-grid one (Figure 3.6). This conclusion does not depend on the time step. From a certain wavenumber on a remarkable behavior, which consists of a constant normalized

[2] More precisely, the 2TL analogue of their scheme B [105].

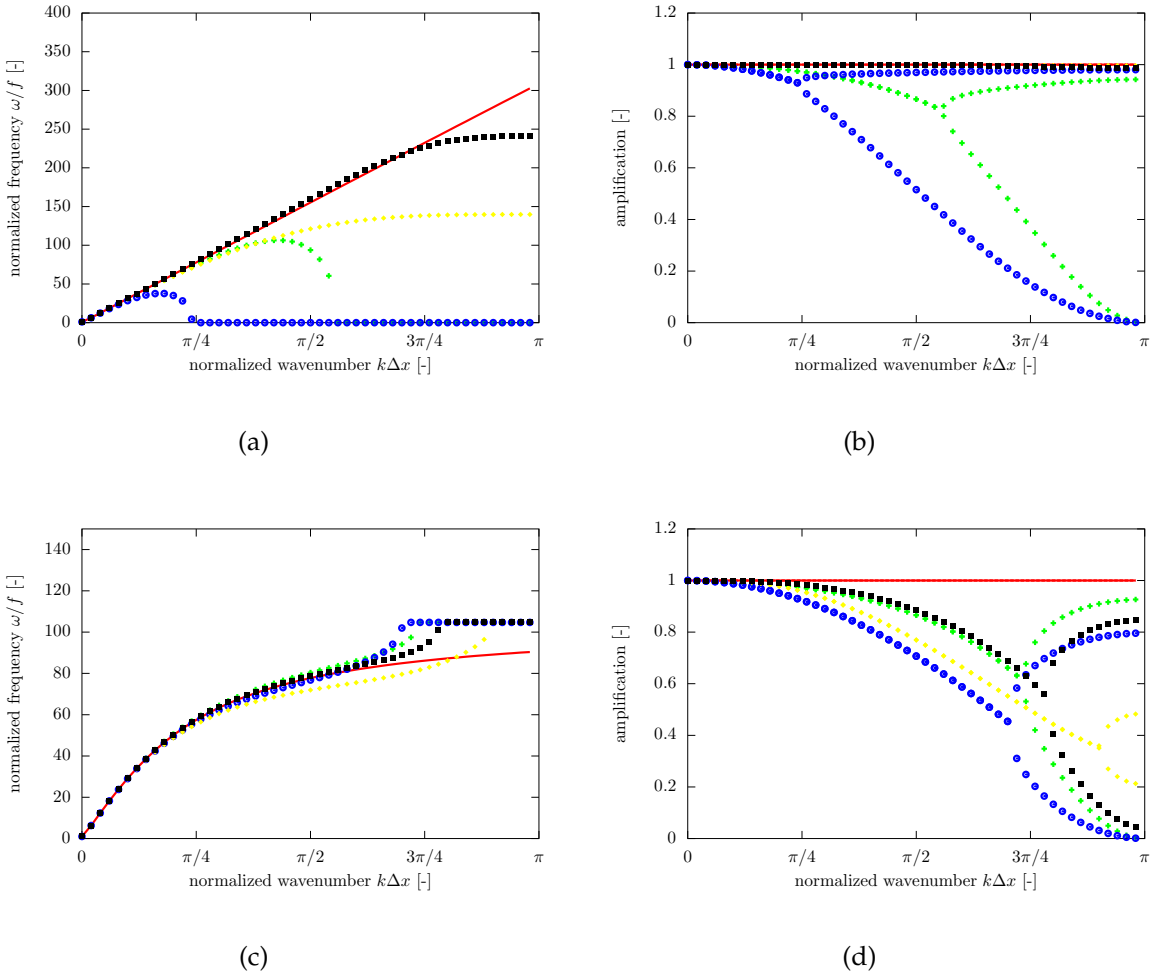


Figure 4.1: IGW normalized frequency (left plots) and amplification factor (right plots) of 2TL SISL Z-grid schemes for $\Delta t = 10\,\mathrm{s}$ (upper plots) and $\Delta t = 300\,\mathrm{s}$ (lower plots). The other parameters are: $\Delta x = 10\,\mathrm{km}$ and $H = 9000\,\mathrm{m}$. The discretization schemes used are: spectral scheme (red cont. line), linear FE ST86 (green crosses), second-order FD ST86 (blue circles), linear FE TS12 (black squares), and second-order FD TS12 (yellow diamonds). In the normalized frequency plots only the IGW propagating in the positive x-direction is shown, whereas in the amplification plot both IGWs are plotted.

frequency and two different values for the amplification, can be noticed for some of the schemes. For these wavenumbers the physical interpretation of two identical IGWs propagating in opposite direction is no longer valid.

- The IGWs are damped, except for the spectral schemes, with the strongest damping for the shortest scale IGWs.
- The qualitative behavior does not depend on whether a second-order FD or linear FE discretization is used.
- One would expect that for small time steps the SI time discretized relations converge towards the Z-grid dispersion relation found with exact time derivation (Figure 3.6). However, Figure 4.1 (a) for a time step $\Delta t = 10\,\mathrm{s}$ shows that this is clearly not the case for the ST86 based schemes.

Is the poor IGW behavior, as seen in Figure 4.1, really problematic? The phase velocity of IGWs is, anyway, strongly slowed down in a SI scheme. Moreover, the poor behavior of the shortest scale IGWs is damped.
There are some good reasons to look for solutions for this dispersion behavior. Contrary to the SI phase velocity retardation, these schemes give rise to IGWs with a group velocity that is negative or even 0, a well-known source for small scale noise. The corresponding damping will maybe limit the propagation of this noise, but it will also distort the energy distribution compared to realistic schemes where the IGWs are neutrally propagating. Moreover, it is generally considered as a bad practice to have compensating errors in an atmospheric model.
The Z-grid approach was constructed precisely to get around the inappropriate dispersion relation of local A-grid approaches. Comparing Figure 3.4 (a) with Figure 4.1 shows that not too much progress is made if no solution can be found for the IGW behavior of Figure 4.1.

4.2.3 Asymmetry in SI Z-grid schemes

By comparing Eq. (4.40) and Eq. (4.41), it can be noticed that there are asymmetries between the implicit matrix $\mathbf{M}_i$ and explicit matrix $\mathbf{M}_e$. The asymmetries are situated in the first row of the matrices, which corresponds to the discretized divergence equation. This causes the poor IGW dispersion of the Z-grid schemes. One can distinguish two different asymmetries:

1. There is an asymmetry between the implicit and explicit second-order derivative operation on the height field. The explicit discretization is based on taking twice a first-order derivative, whereas in the implicit part a second-order derivative discretization is used. This is not identical for most spatial discretization methods:

$\mathcal{P}_{xx} \neq \mathcal{P}_x \mathcal{P}^{-1} \mathcal{P}_x$.

2. Depending on the relation used to retrieve the wind velocities from divergence and vorticity, another asymmetry can be introduced in the divergence equation. This is the case in ST86, because for the PO2 scheme one has: $\mathcal{P}\mathcal{P}_x^* \neq \mathcal{P}_x$. However, the DZ method to restore the wind fields, used in TS12, does not suffer from this problem.

Taking the limit $\Delta t \rightarrow 0$ of Eq. (4.40) and Eq. (4.41) reveals that the second kind of asymmetry makes the scheme even inconsistent. This explains why the dispersion relations found with the ST86 approach are not converging towards the theoretically expected ones for small time steps. The Rossby mode is not altered by these asymmetries, because this mode is mostly controlled by the symmetrically discretized vorticity equation. It was, indeed, noticed in [105] that their scheme A, based on an asymmetrical discretized vorticity equation, results in damping of the slow (Rossby) modes.

4.2.4 IGW analysis of asymmetric discrete divergence equations

In this subsection the influence of temporal asymmetries on the IGW dispersion is investigated for schemes that are based on a SI time discretized divergence equation. It is sufficient to consider only the divergence and continuity equation in the absence of Coriolis terms. Hence the following system,

$$\begin{bmatrix} \mathcal{P}_{x,1} & \frac{g\Delta t}{2}\mathcal{P}_{xx,1} \\ \frac{H\Delta t}{2}\mathcal{P}_{x,1} & \mathcal{P} \end{bmatrix} \begin{bmatrix} u^+ \\ h^+ \end{bmatrix} = \begin{bmatrix} \mathcal{P}_{x,2} & -\frac{g\Delta t}{2}\mathcal{P}_{xx,2} \\ -\frac{H\Delta t}{2}\mathcal{P}_{x,1} & \mathcal{P} \end{bmatrix} \begin{bmatrix} u^0 \\ h^0 \end{bmatrix}, \tag{4.43}$$

is considered where $\mathcal{P}_{x,1}$ and $\mathcal{P}_{x,2}$ represent first-order derivative operators, $\mathcal{P}_{xx,1}$ and $\mathcal{P}_{xx,2}$ second-order derivative operators, and $\mathcal{P}$ the mass operator. A dispersion analysis of the system described by Eq. (4.43) clarifies the impact of asymmetries between implicit and explicit operators.
If one assumes wavelike fields, the operators can be replaced by their corresponding responses and this yields the following amplification matrix:

$$\mathbf{A} = \frac{1}{p_{x,1}\left(p - \beta p_{xx,1}\right)} \begin{bmatrix} p p_{x,2} + \beta p_{x,1} p_{xx,1} & -\frac{g\Delta t}{2} p \left(p_{xx,1} + p_{xx,2}\right) \\ -\frac{H\Delta t}{2} p_{x,1}\left(p_{x,1} + p_{x,2}\right) & p p_{x,1} + \beta p_{x,1} p_{xx,2} \end{bmatrix}, \tag{4.44}$$

with $\beta = gH\left(\frac{\Delta t}{2}\right)^2$. The eigenvalues $\lambda_{\mathbf{A}}$, which determine the dispersion properties of the IGWs, have to obey the following quadratic equation

$$\lambda_{\mathbf{A}}^2 - \mathrm{Tr}(\mathbf{A})\lambda_{\mathbf{A}} + \det(\mathbf{A}) = 0, \tag{4.45}$$

with the trace and the determinant of $\mathbf{A}$ given by

$$\mathrm{Tr}(\mathbf{A}) = \frac{p(p_{x,1}+p_{x,2})+\beta p_{x,1}(p_{xx,1}+p_{xx,2})}{p_{x,1}(p-\beta p_{xx,1})} \tag{4.46}$$

$$\begin{aligned}\det(\mathbf{A}) = &\frac{(pp_{x,2}+\beta p_{x,1}p_{xx,1})(pp_{x,1}+\beta p_{x,1}p_{xx,2})}{p_{x,1}^2(p-\beta p_{xx,1})^2} \\ &-\frac{\beta p_{x,1}p(p_{xx,1}+p_{xx,2})(p_{x,1}+p_{x,2})}{p_{x,1}^2(p-\beta p_{xx,1})^2}.\end{aligned} \tag{4.47}$$

In Table 4.1 the previous expressions together with the discriminant $\Delta(\mathbf{A}) = \mathrm{Tr}^2(\mathbf{A}) - 4\det(\mathbf{A})$ of the quadratic equation (4.45) are calculated for three cases.

Table 4.1: Table containing the trace $\mathrm{Tr}(\mathbf{A})$ and determinant $\det(\mathbf{A})$ of the amplification matrix $\mathbf{A}$ and the discriminant $\Delta(\mathbf{A})$ of the quadratic equation (4.45). Three different cases are distinguished. A more detailed discussion of the consequences for the IGW dispersion behavior is given in the text.

(a) symmetry	$p_{x,1} = p_{x,2} = p_x$	$p_{x,1} = p_{x,2} = p_x$	$p_{x,1} \neq p_{x,2}$
of scheme	$p_{xx,1} = p_{xx,2} = p_{xx}$ (e.g. symmetric scheme of 4.3)	$p_{xx,1} \neq p_{xx,2}$ (asymmetry present in ST86 and TS12)	$p_{xx,1} = p_{xx,2} = p_{xx}$ (asymmetry present in ST86)
$\mathrm{Tr}(\mathbf{A})$	$2\frac{p+\beta p_{xx}}{p-\beta p_{xx}}$	$\frac{2p+\beta(p_{xx,1}+p_{xx,2})}{p-\beta p_{xx,1}}$	$\frac{p(p_{x,1}+p_{x,2})+2\beta p_{x,1}p_{xx}}{p_{x,1}(p-\beta p_{xx,1})}$
$\det(\mathbf{A})$	1	$\frac{(p+\beta p_{xx,1})(p+\beta p_{xx,2})-2\beta p(p_{xx,1}+p_{xx,2})}{(p-\beta p_{xx,1})^2}$	$\frac{(pp_{x,2}+\beta p_{x,1}p_{xx})(pp_{x,1}+\beta p_{x,1}p_{xx})-2\beta pp_{x,1}p_{xx}(p_{x,1}+p_{x,2})}{p_{x,1}^2(p-\beta p_{xx})^2}$
$\Delta(\mathbf{A})$	$16\frac{\beta p_{xx}p}{(p-\beta p_{xx})^2}$	$\frac{\beta^2\left[(p_{xx,1}+p_{xx,2})^2-4p_{xx,1}p_{xx,2}\right]+8\beta p(p_{xx,1}+p_{xx,2})}{(p-\beta p_{xx,1})^2}$	$\frac{s^2(p_{x,1}-p_{x,2})^2+8\beta pp_{x,1}p_{xx}(p_{x,1}+p_{x,2})}{p_{x,1}^2(p-\beta p_{xx})^2}$
impact on IGWs	2 neutral IGWs	$\Delta t << $: 2 dampened IGWs $\Delta t >> $: 2 dampened IGWs for $k <<$ and unphysical solutions for $k >>$	$\Delta t << $: 2 dampened IGWs for $k <<$ and unphysical solutions for $k >$ $\Delta t >> $: 2 dampened IGWs for $k <<$ and unphysical solutions for $k >>$

Assume first a completely symmetric scheme. This yields: $p_{x,1} = p_{x,2} = p_x$ and $p_{xx,1} = p_{xx,2} = p_{xx}$ (first column in Table 4.1). Because the responses of second-order derivative operators are always negative for spectral, linear FE, and second-order FD schemes (see Table 3.1), the discriminant corresponding to this kind of scheme will be negative independent of the time step or the wavenumber. This corresponds to two complex conjugate eigenvalues: $\lambda^*_{\mathbf{A},1} = \lambda_{\mathbf{A},2}$. The product of the two roots of (4.45) must be equal to $\det(\mathbf{A})$, in this case 1, implying that both IGW solutions are neutral for the symmetric scheme. To summarize, the symmetric scheme will give rise to two neutral IGWs propagating with the same phase velocity in the positive and negative x-direction, in agreement with the physical behavior of IGWs.
In case a temporal asymmetry is introduced, the IGW properties change drastically. Assume we have $p_{x,1} = p_{x,2} = p_x$ but two different second-order derivative operators $p_{xx,1} \neq p_{xx,2}$. The discriminant consists of two terms, a negative one $8\beta p\left(p_{xx,1} + p_{xx,2}\right)$ and a positive one $\beta^2\left[\left(p_{xx,1} + p_{xx,2}\right)^2 - 4p_{xx,1}p_{xx,2}\right]$. This means that β (containing the time step) and the wavenumber will determine the IGW behavior: for short time steps and short wavenumbers we will find the expected IGW behavior (corresponding to a negative discriminant and thus two complex conjugate solutions), longer time steps will lead to an increasing number of IGWs with zero group velocity (corresponding to a positive discriminant and thus two real eigenvalues). The latter behavior is no longer related to analytical IGWs and is therefore qualified as unphysical. The expression of $\det(\mathbf{A})$ shows we will no longer have two neutral solutions. Whether the waves are damped or amplified depends on the exact form of $p_{xx,1}$ and $p_{xx,2}$.
The third column of Table 4.1 describes a scheme with a temporal asymmetry for the first-order derivative operation $p_{x,1} \neq p_{x,2}$. This kind of scheme will have a positive discriminant and thus unphysical IG modes even in the limit of very small time steps ($\beta \to 0$).
This analysis confirms the results of Figure 4.1 remembering that ST86 contains both types of asymmetries, whereas TS12 only has $p_{xx,1} \neq p_{xx,2}$.

4.3 Symmetric SISL Z-grid schemes

The previous section makes clear that one should avoid asymmetries between the operations at time t^0 and t^+. In what follows, symmetric SISL Z-grid schemes are proposed and tested by using a toy model based on the SWE.

4.3.1 Formulation of symmetric 2TL SISL Z-grid schemes

1D symmetric 2TL SISL Z-grid schemes

Let us retake the approach used in 4.2.1 but this time with a vorticity equation that is constructed from the time discretized momentum equations, as it is done in ALADIN (Eq. 4.8). This scheme can be written in the following matrix form:

$$\begin{aligned}
&\begin{bmatrix} \mathcal{P} & 0 & \beta_f \frac{g\Delta t}{2}\mathcal{P}_{xx} \\ \frac{f\Delta t}{2}\mathcal{P} & \mathcal{P} & 0 \\ \frac{H\Delta t}{2}\mathcal{P} & 0 & \mathcal{P} \end{bmatrix} \begin{bmatrix} \mathcal{P}_x^* & 0 & 0 \\ 0 & \mathcal{P}_x^* & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u^+ \\ v^+ \\ h^+ \end{bmatrix} \\
&= \begin{bmatrix} \beta_f \mathcal{P}_{x,1} & \beta_f \frac{f\Delta t}{2}\mathcal{P}_{x,1} & 0 \\ 0 & \mathcal{P}_{x,1} & 0 \\ 0 & 0 & \mathcal{P} \end{bmatrix} \begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{g\Delta t}{2}\mathcal{P}_{x,2} \\ -\frac{f\Delta t}{2} & 1 & 0 \\ -\frac{H\Delta t}{2}\mathcal{P}_x^* & 0 & 1 \end{bmatrix} \begin{bmatrix} u^0 \\ v^0 \\ h^0 \end{bmatrix},
\end{aligned} \tag{4.48}$$

where $\mathcal{P}_x^*$ defines how to retrieve the wind components from D and ζ. In the explicit calculations two different derivative operators $\mathcal{P}_{x,1}$ and $\mathcal{P}_{x,2}$ are introduced. They should be specified in such a way that the symmetry constraints are satisfied. Remark that the operators in the implicit part are left untouched. Modifying them such that the scheme is symmetric would result in an A-grid like scheme, as will be explained in section 4.5.

Given $\mathcal{P}_x^*$, there is in Eq. (4.48) one unique choice for $\mathcal{P}_{x,1}$ and $\mathcal{P}_{x,2}$ that results in symmetric equations:

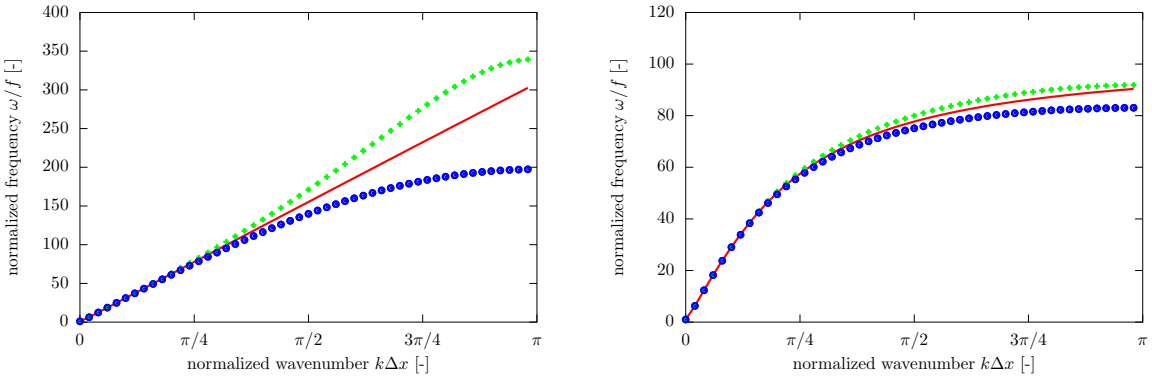
$$\mathcal{P}_{x,1} = \mathcal{P}\mathcal{P}_x^* \tag{4.49}$$

$$\mathcal{P}_{x,2} = \mathcal{P}_{x,1}^{-1}\mathcal{P}_{xx} = \left(\mathcal{P}_x^*\right)^{-1}\mathcal{P}^{-1}\mathcal{P}_{xx}. \tag{4.50}$$

Given the three possibilities for the wind reconstruction, different symmetric 2TL SISL Z-grid schemes, which are listed in Table 4.2, are constructed. These schemes have different amplification matrices. However, it was verified that they result in the same IGW dispersion behavior, which is illustrated in Figure 4.2 for a second-order FD and a linear FE spatial discretization method and for two choices of the time step. As expected, the IGWs are neutral (not shown here) and their phase velocity is close to the spectral one. This agrees with the excellent dispersion found in 3.3.3 for the Z-grid approach.

In practice, these schemes are solved in the following way:

1. Evaluate the right hand sides R_u, R_v, and R_h, which are given by (4.27), (4.28), and (4.29) with $\mathcal{P}_x$ replaced by $\mathcal{P}\mathcal{P}_{x,2}$ in the evaluation of the height gradient.



(a) (b)

Figure 4.2: Normalized IGW frequency for symmetric 2TL SISL Z-grid schemes (4.48) combined with a second-order FD (blue circles) and a linear FE (green crosses) horizontal spatial discretization scheme. The parameters are identical to the ones in Figure 4.1: $\Delta t = 10\,\mathrm{s}$ (left) and $\Delta t = 300\,\mathrm{s}$ (right). These plots are independent of the choice of $\mathcal{P}_x^*$. The spectral dispersion relation (red cont. line) is plotted for reference.

Table 4.2: The first-order derivative operators $\mathcal{P}_{x,1}$ and $\mathcal{P}_{x,2}$ needed for the three different wind retrieval schemes (DZ, PO1, and PO2 reconstruction) to end up with a symmetric Z-grid scheme.

	DZ	PO1	PO2
$\mathcal{P}_x^*$	$\mathcal{P}^{-1}\mathcal{P}_x$	$\mathcal{P}_x\left(\mathcal{P}_x^2\right)^{-1}\mathcal{P}_{xx}$	$\mathcal{P}^{-1}\mathcal{P}_x\left(\mathcal{P}_x^2\right)^{-1}\mathcal{P}\mathcal{P}_{xx}$
$\mathcal{P}_{x,1}$	$\mathcal{P}_x$	$\mathcal{P}\mathcal{P}_x\left(\mathcal{P}_x^2\right)^{-1}\mathcal{P}_{xx}$	$\mathcal{P}_x\left(\mathcal{P}_x^2\right)^{-1}\mathcal{P}\mathcal{P}_{xx}$
$\mathcal{P}_{x,2}$	$\mathcal{P}_x\left(\mathcal{P}_x^2\right)^{-1}\mathcal{P}_{xx}$	$\mathcal{P}_{xx}^{-1}\mathcal{P}_x\mathcal{P}^{-1}\mathcal{P}_{xx}$	$\mathcal{P}_{xx}^{-1}\mathcal{P}^{-1}\mathcal{P}_x\mathcal{P}_{xx}$

2. Solve the Helmholtz equation (4.31) with $\mathcal{P}_x$ replaced by $\mathcal{P}_{x,1}$. Afterwards, one calculates divergence D^+ and ζ^+ easily by solving:

$$\mathcal{P}h^+ + \frac{H\Delta t}{2}\mathcal{P}D^+ = \mathcal{P}R_h \tag{4.51}$$

$$\mathcal{P}\zeta^+ + \frac{f\Delta t}{2}\mathcal{P}D^+ = \mathcal{P}_{x,1}R_v. \tag{4.52}$$

3. Retrieve the wind components u^+ and v^+ by solving:

$$u^+ = (\mathcal{P}_x^*)^{-1} D^+ \tag{4.53}$$

$$v^+ = (\mathcal{P}_x^*)^{-1} \zeta^+. \tag{4.54}$$

This time step organization is very similar to the time stepping of the asymmetric scheme described in 4.2.1. However, a price is paid to restore symmetry. Operator $\mathcal{P}_x$ is replaced by more complex differential operators in the right hand sides of the momentum equation R_u, the Helmholtz equation, and the vorticity equation.

2D symmetric 2TL SISL Z-grid schemes

Let us repeat the previous for the 2D linearized SWE. However, in order not to overload the expressions, wavelike fields are assumed. In this way the operators can be replaced by their responses, which commute and, therefore, simplify the expressions. The 2D matrix formulation of the linearized SWE then takes the following form:

$$\begin{bmatrix} q & -\frac{f\Delta t}{2}q & \frac{g\Delta t}{2}\left(q_{xx}+q_{yy}\right) \\ \frac{f\Delta t}{2}q & q & 0 \\ \frac{H\Delta t}{2}q & 0 & q \end{bmatrix} \begin{bmatrix} q_x^* & q_y^* & 0 \\ -q_y^* & q_x^* & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u^+ \\ v^+ \\ h^+ \end{bmatrix}$$

$$= \begin{bmatrix} q_{x,1} & q_{y,1} & 0 \\ -q_{y,1} & q_{x,1} & 0 \\ 0 & 0 & q \end{bmatrix} \begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{g\Delta t}{2}q_{x,2} \\ -\frac{f\Delta t}{2} & 1 & -\frac{g\Delta t}{2}q_{y,2} \\ -\frac{H\Delta t}{2}q_x^* & -\frac{H\Delta t}{2}q_y^* & 1 \end{bmatrix} \begin{bmatrix} u^0 \\ v^0 \\ h^0 \end{bmatrix}, \tag{4.55}$$

where the 2D responses, denoted by e.g., q, depend on both k and l. Contrary to Eq. (4.48), vorticity is not yet eliminated from the discrete divergence equation. This makes it easier to identify the form of the operators needed to end up with a symmetric scheme. Table 4.3 gives the expression of the different 2D responses needed to end up with a symmetric scheme. Due to commutativity of the responses, the PO1 and PO2 reconstruction schemes are identical.

Table 4.3: The first-order derivative responses $q_{x,1}$ and $q_{x,2}$ needed for the different wind retrieval schemes (DZ and PO reconstruction) to end up with a 2D symmetric Z-grid scheme. Similar expressions are found for the first-order y-derivative responses.

	DZ scheme	PO1 = PO2 scheme
q_x^*	$\frac{q_x}{q}$	$q_x \frac{q_{xx}+q_{yy}}{q_x^2+q_y^2}$
$q_{x,1}$	q_x	$qq_x \frac{q_{xx}+q_{yy}}{q_x^2+q_y^2}$
$q_{x,2}$	$q_x \frac{q_{xx}+q_{yy}}{q_x^2+q_y^2}$	$\frac{q_x}{q}$

4.3.2 Numerical SWE toy model tests

In this subsection the IGW dispersion of symmetric and asymmetric SI Z-grid schemes is discussed on the basis of toy model tests of the 1D and 2D linearized SWE. In what follows the equations are formulated in terms of geopotential ϕ instead of height h. Some trivial modifications are sufficient to reformulate the schemes derived in the previous sections in terms of geopotential.

1D linearized SWE tests

In [111] an initial state

$$\phi_0(x) = A_\phi \cos\left(\frac{2\pi x}{\lambda}\right) e^{-\left(\frac{x-x_0}{L_\sigma}\right)^2} \tag{4.56}$$

$$u_0(x) = 0, \tag{4.57}$$

consisting of two symmetrically propagating IGWs (one leftward and one rightward), was assumed to study the dispersion properties of mixed FE schemes numerically. In this publication the Coriolis force was not taken into account ($f = 0$).
Hereafter, this IGW propagation test is repeated to study the differences between symmetric and asymmetric SI Z-grid schemes. The test set-up consists of IGWs with central wavelength $\lambda = 4\Delta x$, situated initially in the middle of the domain ($x_0 = L/2$) with amplitude $A_\phi = 10\,\mathrm{m^2\ s^{-2}}$ and $L_\sigma = L/20$. The domain consists of 400 grid points with grid spacing $\Delta x = 10\,\mathrm{km}$. The time step used for the SI integration is 10 s and the reference geopotential $\Phi = 9000\,\mathrm{m^2\ s^{-2}}$.
Figure 4.3 shows the integration in time, starting from the initial state (4.56) - (4.57), of the following schemes:

- a spectral SI scheme (symmetric or asymmetric does not matter if spectral spatial discretization is used) (Figure 4.3 (a) - (b)),
- a symmetric linear FE Z-grid scheme based on DZ reconstruction (Figure 4.3 (c) -

(d)), and

- a scheme based on the method employed in TS12 (asymmetric for the second-order derivative of the geopotential field) with a linear FE scheme used for the spatial discretization (Figure 4.3 (e) - (f)).

The left column of Figure 4.3 shows snapshots of the geopotential field ϕ at different times. A space-time diagram, which shows the evolution of the geopotential, is plotted in the right column.
The integrations based on the spectral and symmetric Z-grid schemes show the expected IGW propagation. The larger group velocity of the latter scheme corresponds to the differences in the slope of their phase velocities around $k\Delta x = \pi/2$ in Figure 4.2. However, the asymmetric scheme slowly damps the IGWs, as can be seen in Figure 4.3 (e) and (f). Use of the asymmetric ST86 scheme (containing an extra asymmetry apart from the one already present in the TS12 scheme) results in very strong damping of the IGWs. It was verified (not shown in Figure 4.3) that after only some time steps the fields are damped completely with this approach. We conclude that the numerical results agree well with the analytical study presented in the previous sections.

The 1D geostrophic adjustment test, decribed in [50], forms another interesting numerical test. This time a Heaviside function is applied as the initial perturbation around the reference geopotential Φ. Again there is no initial velocity. This strongly unbalanced initial state will evolve towards geostrophic balance by propagating IGWs away from the barrier. In this test Coriolis terms and thus the meridional velocity v are included. Figure 4.4 is a numerical illustration of this geostrophic adjustment process and shows the geopotential field after 20 time integrations ($\Delta t = 300\,\mathrm{s}$) for 4 different Z-grid schemes. It can be seen that the shortest scale IGWs are strongly suppressed by the asymmetric Z-grid schemes. As expected, their representation of the longer IGWs is similar to the one found with the spectral and symmetric Z-grid scheme. However, damping of the short waves will change the energy distribution. The symmetric Z-grid scheme does not suffer from this problem and results (apart from a phase speed difference) in wave propagation similar to the spectral scheme for the whole wavelength range.

2D linearized SWE tests

The 2D test applied in [71] to compare 2D mixed FE methods with and without mass lumping is used here in the context of (a)symmetric 2D Z-grid schemes. The initial

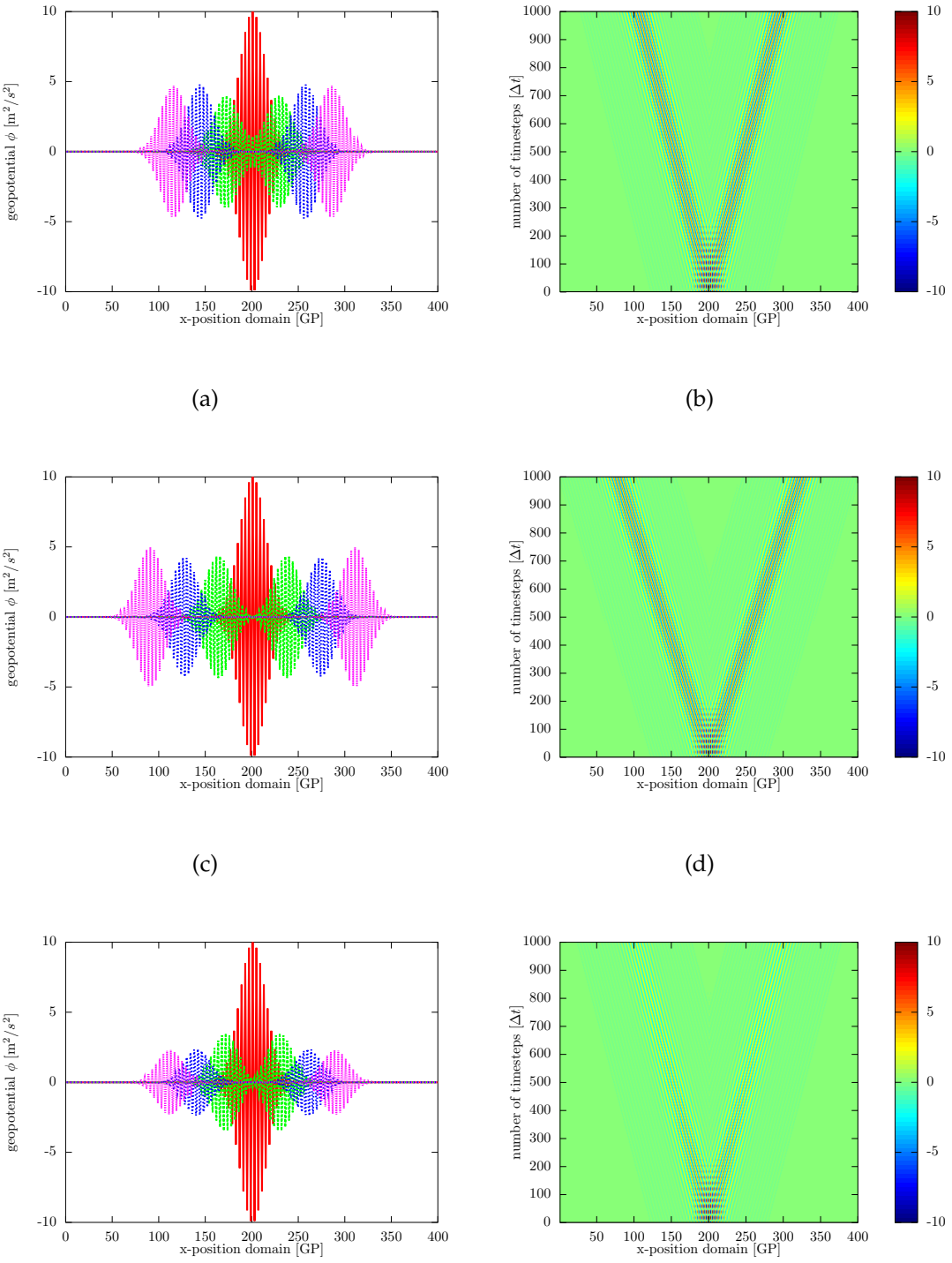
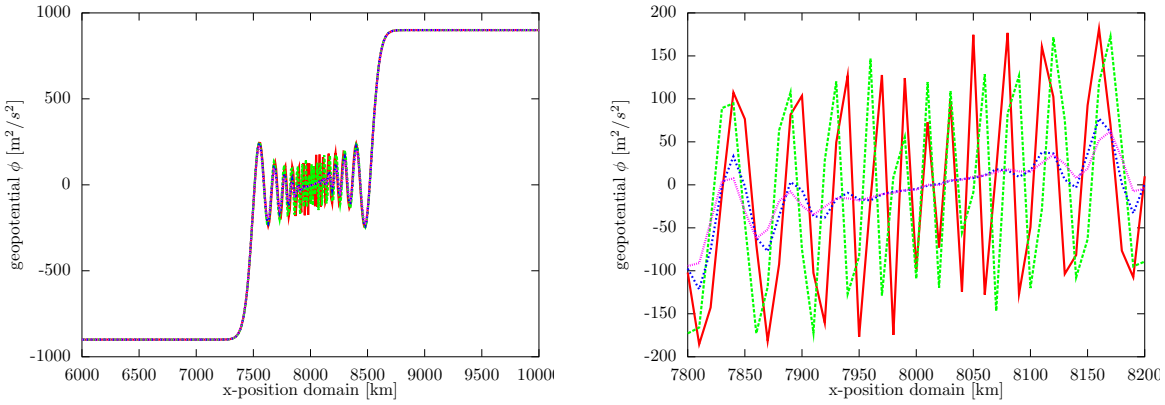


Figure 4.3: 1D IGW propagation test similar to the one in [111]. From top to bottom the results are shown for a spectral scheme, a linear FE symmetric scheme (DZ reconstruction), and a linear FE TS12 scheme. The left column show the geopotential field ϕ at the initial time (red), after 300 (green), 600 (blue), and 900 (magenta) time integrations. The right column displays the evolution of the geopotential field in a space-time diagram. The parameter values of the test are: $\Delta x = 10\,\text{km}$, $\Delta t = 10\,\text{s}$, and $\Phi = 9000\,\text{m}^2\,\text{s}^{-2}$.



(a) (b)

Figure 4.4: The geopotential field after 20 time steps in a 1D geostrophic adjustment test with parameters: $\Delta t = 300\,\mathrm{s}$, $\Delta x = 10\,\mathrm{km}$, $\Phi = 9000\,\mathrm{m}^2\,\mathrm{s}^{-2}$, and the height of the step is 1800 $\mathrm{m}^2\,\mathrm{s}^{-2}$. The right plot is a zoom in of the left plot around the location of the barrier. The spectral (red) and the linear FE symmetric scheme based on DZ reconstruction (green) show propagating IGWs, whereas the ST86 scheme (magenta) and the linear FE TS12 scheme (blue) damp all short scale IGWs.

state consists of two IGWs and a Rossby wave and is given by:

$$\phi_0(x,y) = A_\phi \cos\left(\frac{2\pi x}{\lambda_x}\right)\cos\left(\frac{2\pi y}{\lambda_y}\right) e^{-\left(\frac{x-x_0}{L_{x,\sigma}}\right)^2} e^{-\left(\frac{y-y_0}{L_{y,\sigma}}\right)^2}, \tag{4.58}$$

$$u_0(x,y) = 0, \text{ and} \tag{4.59}$$

$$v_0(x,y) = 0. \tag{4.60}$$

Analytical time integration of this initial state gives rise to IGWs propagating along the diagonals of the (x,y)-plane. In the center of the domain (around (x_0,y_0)) a stationary Rossby mode is expected. It can be seen in Figure 4.5, which shows the geopotential field 50 h after the start of the run, that both the spectral and the symmetric scheme demonstrate this behavior. It does not come as a surprise that the IGWs are damped partly and completely if the TS12 and ST86 scheme are used, respectively.

It is found that every discretization method results in a different Rossby mode (in the center of the domain). At first sight this is a surprising result because the stationary character of this mode is guaranteed by all schemes. The differences are indeed not related to the propagation properties of the scheme, but they are due to different projections of the initial state onto the discretized Rossby and IGW modes of the schemes.

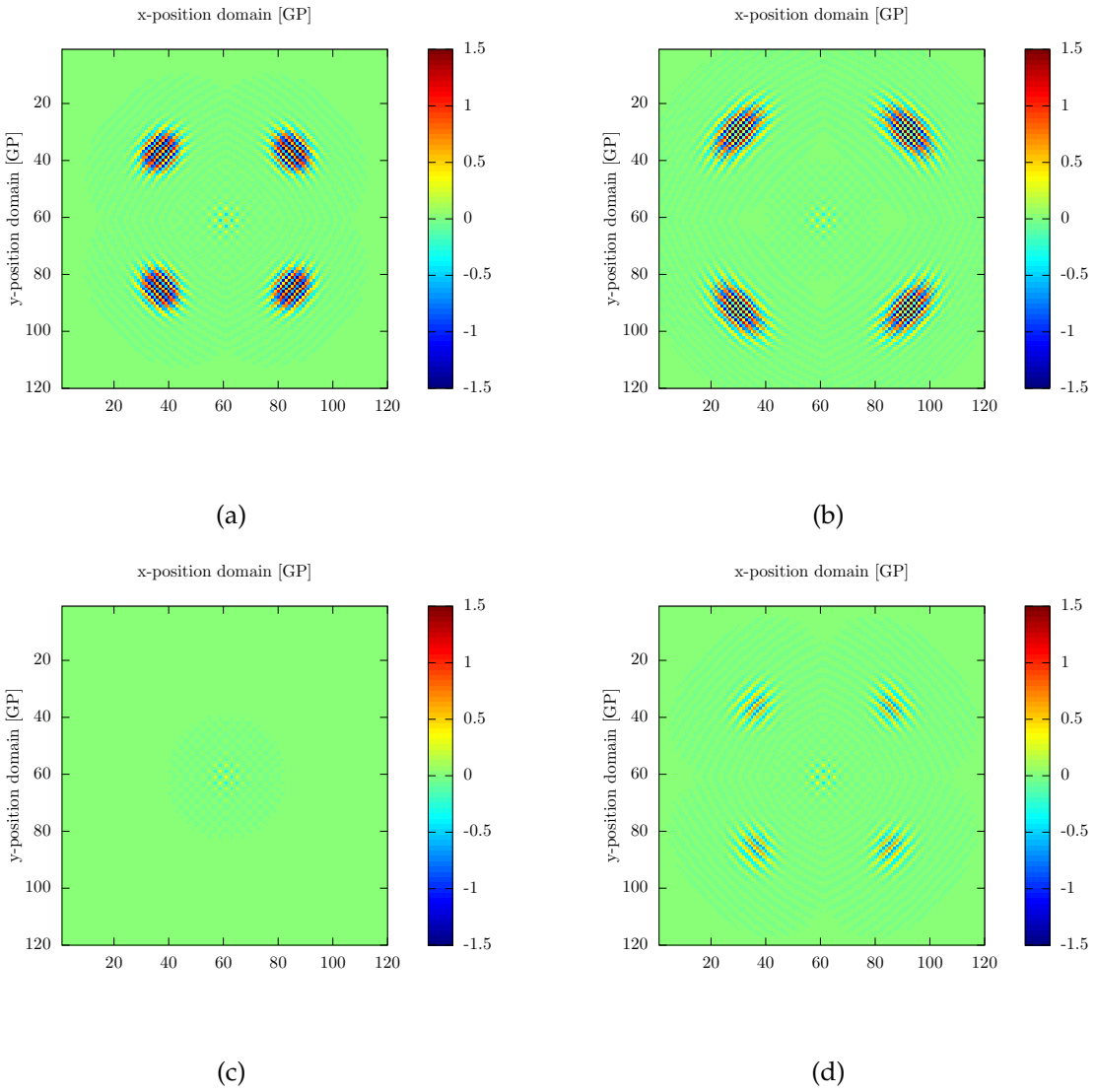
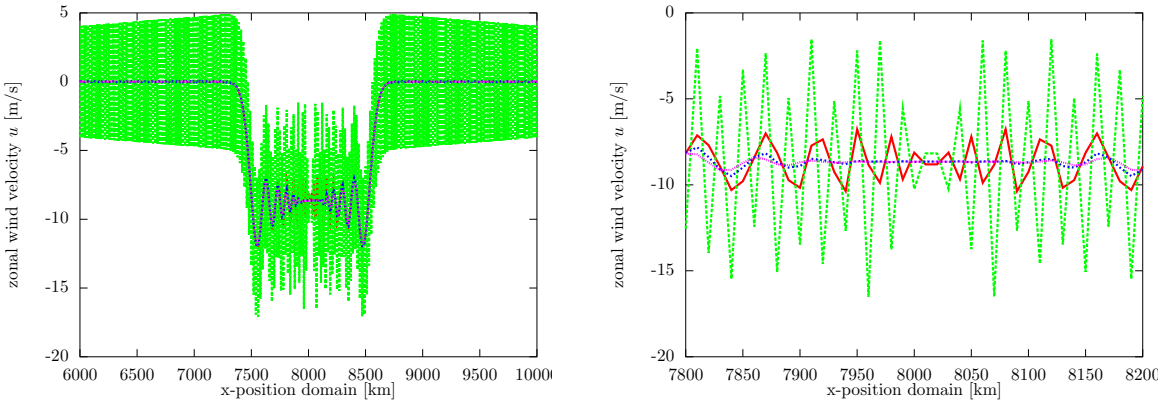


Figure 4.5: 2D test similar to the one in Figure 8 of [71]. Plot of the geopotential ϕ after 50 h: spectral scheme (a), linear FE symmetric scheme based on DZ reconstruction (b), linear FE ST86 scheme (c), and linear FE TS12 scheme (d). The used parameter values are: $\Delta x = \Delta y = 50\,\text{km}$, 120 grid points in x- and y-direction, $\Delta t = 800\,\text{s}$, $\Phi = 100\,\text{m}^2\,\text{s}^{-2}$, $A_\Phi = 10\,\text{m}^2\,\text{s}^{-2}$, $\lambda_x = \lambda_y = 4\Delta x$ and $L_{x,\sigma} = L_{y,\sigma} = 6\Delta x$.



(a) (b)

Figure 4.6: The zonal wind velocity u after 20 time steps in a 1D geostrophic adjustment test with parameters identical to the one in Figure 4.4.

4.4 IGW eigenmodes of Z-grid schemes

The Z-grid SISL scheme seems to be an elegant way to conserve appropriate geostrophic adjustment in combination with local discretization schemes on an unstaggered grid on the condition that the symmetry between explicit and implicit operators is guaranteed. However, reproducing Figure 4.4 for the zonal velocity u gives an unexpected result as can be seen in Figure 4.6. The wind field shows a strong, short scale noise component on top of the expected field if the symmetric SISL Z-grid is used. This is clearly unphysical and thus unwanted. This behavior comes as a surprise as:

- there is no problem with the geopotential field for this scheme, as can be seen in Figure 4.4, and
- the dispersion analysis, which describes the behavior of the waves, did not give any clue for this.

Despite being visible only for the symmetric scheme in Figure 4.6, the problem is generally valid for local discretization schemes combined with a divergence and vorticity formulation. However, in the asymmetric Z-grid schemes the noise is masked by the damping.
In what follows, we present evidence that this behavior is related to the *eigenvectors* or *eigenmodes* of the Z-grid scheme.

Consider a linear time-continuous system

$$\frac{\partial}{\partial t}\mathbf{X} = \mathbf{L}\mathbf{X}, \tag{4.61}$$

where the amplification matrix $\mathbf{L}$ determines the evolution in time of the state vector $\mathbf{X}$. In general (if $\mathbf{L}$ contains off-diagonal elements) the different components of the state vector $\mathbf{X}$ are coupled. If the matrix $\mathbf{L}$ has eigenvectors $\mathbf{v}_i$ and eigenvalues λ_i, one has by definition:

$$\mathbf{L}\mathbf{V} = \mathbf{V}\mathbf{\Lambda},$$

where the columns of $\mathbf{V}$ contain the eigenvectors and $\mathbf{\Lambda}$ is the diagonal matrix containing the eigenvalues λ_i. The previous permits to rewrite Eq. (4.61) into a diagonal or decoupled form:

$$\frac{\partial}{\partial t}\mathbf{W} = \mathbf{\Lambda}\mathbf{W} \quad \text{with} \quad \mathbf{W} = \mathbf{V}^{-1}\mathbf{X}. \tag{4.62}$$

This reveals two important properties of $\mathbf{L}$:

1. $\mathbf{W}$ is the projection of the physical state vector $\mathbf{X}$ onto the basis spanned by the eigenvectors or, put differently, the eigenvectors of $\mathbf{L}$ determine the uncoupled eigenmodes of the physical problem.
2. The eigenvalues of $\mathbf{L}$ determine the propagation (damping, phase velocity,...) of the waves.

In literature, numerous eigenvalue studies were done to study the impact of discretization choices. However, this is no sufficient condition to obtain an appropriate scheme. Even a numerical scheme possessing the analytical eigenvalues can give completely incorrect results if the eigenvectors are erroneous. Both eigenvalues and eigenvectors are important if one wants to study the properties of a scheme. In what follows, this is illustrated by relating the noisy u-field of the adjustment test of Figure 4.6 to the eigenvectors of the Z-grid approach.

The 1D SWE suffice to analyze the eigenvector problem of the Z-grid. Following the form of Eq. (4.61) and assuming wave solutions, the space discretized equations formulated in terms of momentum write as:

$$\frac{\partial}{\partial t}\begin{bmatrix} u \\ \phi \end{bmatrix} = \mathbf{L_{AC}}\begin{bmatrix} u \\ \phi \end{bmatrix} = \begin{bmatrix} 0 & -p_x \\ -\Phi p_x & 0 \end{bmatrix}\begin{bmatrix} u \\ \phi \end{bmatrix}, \tag{4.63}$$

with p_x the earlier introduced 1D response of the first-order derivative operator. Eq. (4.63) forms the starting point for the A-grid and staggered grid (e.g., C-grid) discretizations.

The 1D SWE formulated in terms of divergence yield

$$\frac{\partial}{\partial t}\begin{bmatrix} D \\ \phi \end{bmatrix} = \begin{bmatrix} 0 & -p_{xx} \\ -\Phi p & 0 \end{bmatrix}\begin{bmatrix} D \\ \phi \end{bmatrix}.$$

Using the relation between divergence and zonal wind $D = p_x^* u$, one rewrites the previous into a form with the same state vector as Eq. (4.63)

$$\frac{\partial}{\partial t}\begin{bmatrix} u \\ \phi \end{bmatrix} = \mathbf{L_Z}\begin{bmatrix} u \\ \phi \end{bmatrix} = \begin{bmatrix} 0 & -\frac{p_{xx}}{p_x^*} \\ -\Phi p p_x^* & 0 \end{bmatrix}\begin{bmatrix} u \\ \phi \end{bmatrix}. \tag{4.64}$$

Calculating the eigenvalues and eigenvectors of Eq. (4.63) and Eq. (4.64) results into

$$\mathbf{L}_{AC} = \begin{bmatrix} 0 & -p_x \\ -\Phi p_x & 0 \end{bmatrix} \longrightarrow \lambda_{AC}^{\pm} = \pm\sqrt{\Phi} p_x \text{ and } \mathbf{V}_{AC} = \begin{bmatrix} 1 & 1 \\ \sqrt{\Phi} & -\sqrt{\Phi} \end{bmatrix}$$

$$\mathbf{L}_Z = \begin{bmatrix} 0 & -\frac{p_{xx}}{p_x^*} \\ -\Phi p p_x^* & 0 \end{bmatrix} \longrightarrow \lambda_Z^{\pm} = \pm\sqrt{\Phi}\sqrt{p p_{xx}} \text{ and } \mathbf{V}_Z = \begin{bmatrix} \sqrt{p_{xx}} & \sqrt{p_{xx}} \\ \sqrt{\Phi}\sqrt{p} p_x^* & -\sqrt{\Phi}\sqrt{p} p_x^* \end{bmatrix}.$$

For the momentum formulation (4.63) two main conclusions can be drawn. First, the eigenvectors do not depend on the responses. Put differently, using FD, FE, or staggered FD does not impact the form of the eigenvectors although this choice does impact the eigenvalues and thus the propagation properties of the waves. Moreover, the eigenvectors of the momentum formulation are identical to the analytical ones. Secondly, the amplitude of the geopotential component of the eigenvectors is much larger than the amplitude of u.
However, the eigenvectors of the Z-grid approach (4.64) do depend on the wavenumber and the expression for the responses, except for the spectral approach where $\sqrt{p_{xx}} = p_x$, $p_x^* = p_x$, and $p = 1$ and one obtains again the analytical form. In the long wave limit, $k \to 0$, we obtain eigenvectors converging to the analytical ones. However, for the shortest waves an unexpected behavior can be noticed. If one uses for example DZ reconstruction in combination with second-order FD, one gets for the eigenvectors

$$\mathbf{V}_Z^{DZ} = \begin{bmatrix} \sqrt{2 - 2\cos(k\Delta x)} & \sqrt{2 - 2\cos(k\Delta x)} \\ \sqrt{\Phi}\sin(k\Delta x) & -\sqrt{\Phi}\sin(k\Delta x) \end{bmatrix},$$

which means that for $k\Delta x \to \pi$ the ϕ-components of the eigenvectors approach 0. Thus, for short waves the decoupling of the initial state into two waves is dominated

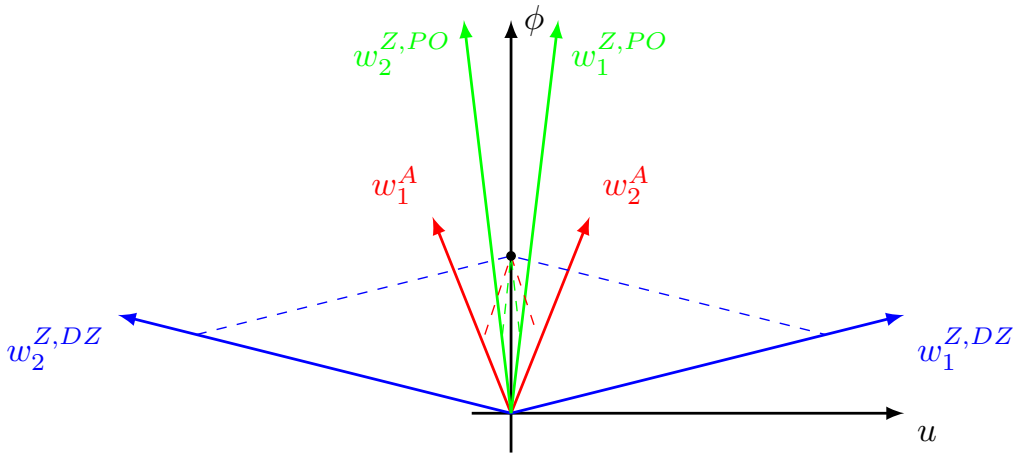


Figure 4.7: Projection of a short wave component of an initial state in rest (black dot on the y-axis) onto the two eigenvectors. The decoupling is different for the A-grid scheme (red), the DZ-based Z-grid scheme (blue), and the PO-based Z-grid scheme (green).

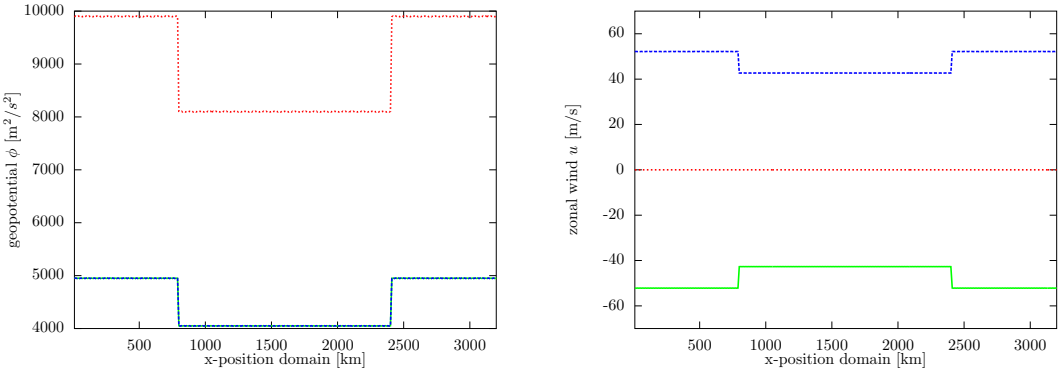
by u instead of ϕ. The reverse is seen if PO reconstruction is used. We then have

$$\mathbf{V}_Z^{PO} = \begin{bmatrix} \sin(k\Delta x) & \sin(k\Delta x) \\ \sqrt{\Phi}\sqrt{2-2\cos(k\Delta x)} & -\sqrt{\Phi}\sqrt{2-2\cos(k\Delta x)} \end{bmatrix},$$

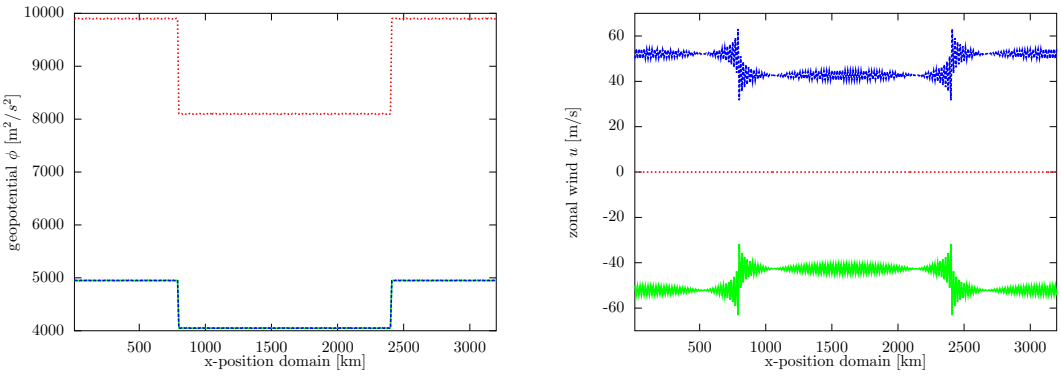
which means that in the short scale limit the rate between the ϕ-component and the u-component becomes infinitely large. The decoupling of a short scale wave for different schemes is illustrated in Figure 4.7.

The previous analysis was undertaken for a time continuous problem. Let us now verify with a numerical experiment the impact of this eigenvector problem for SI time discretized schemes.

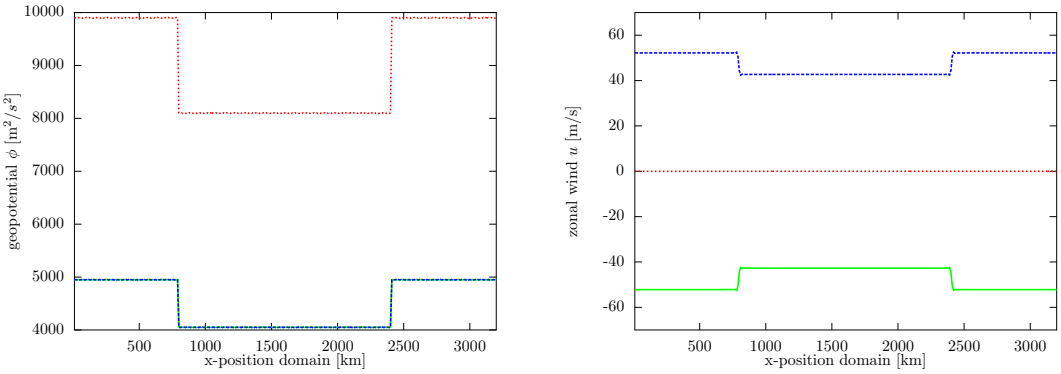
We retake the 1D geostrophic adjustment test introduced in 4.3.2, but the Coriolis force will be omitted this time. The initial state is projected onto the two IGW eigenmodes of the scheme. The coefficients of the two modi should be identical in order to end up with an initial state in rest or, put graphically, to end up on the ϕ-axis ($u = 0$) in Figure 4.7. Figure 4.8 shows this projection of the initial state for the spectral and symmetric Z-grid schemes discretized with second-order FD. While the total field (in red) is identical, the decomposition into the two eigenstates (in blue and green) differs between the two approaches. As explained earlier, the eigenmodes of the Z-grid scheme with DZ-reconstruction have a considerable u-component for the shortest waves and



(a) (b)

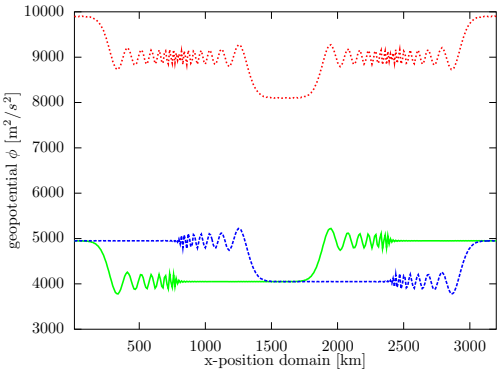


(c) (d)

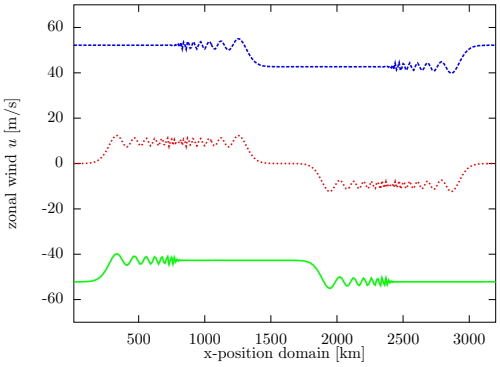


(e) (f)

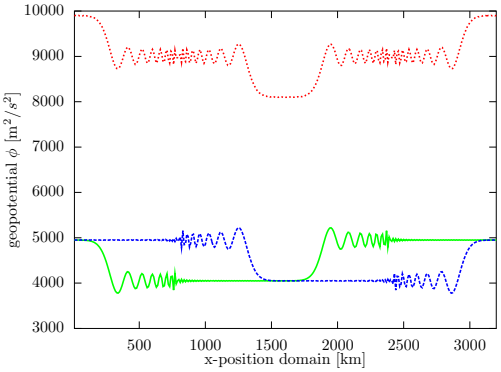
Figure 4.8: Plot of the geopotential $\phi + \Phi$ and zonal wind u of a geostrophic adjustment test at the initial time: spectral SI scheme (a, b), symmetric second-order FD SI Z-grid scheme with DZ reconstruction (c, d), and symmetric second-order FD SI Z-grid scheme with PO reconstruction (e, f). The green and blue line represent the projection of the initial state onto the IGW eigenmodes, the red line is the total field. The settings are: $\Delta x = 10\,\text{km}$, $\Delta t = 300\,\text{s}$, and $\Phi = 9000\,\text{m}^2\,\text{s}^{-2}$.



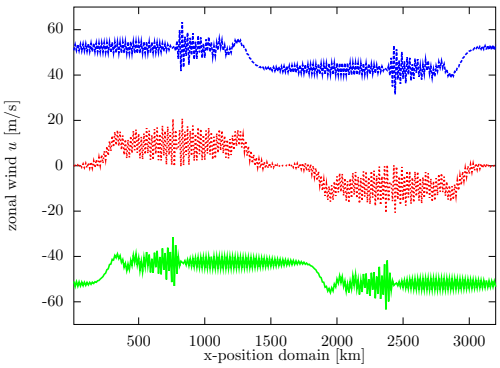
(a)



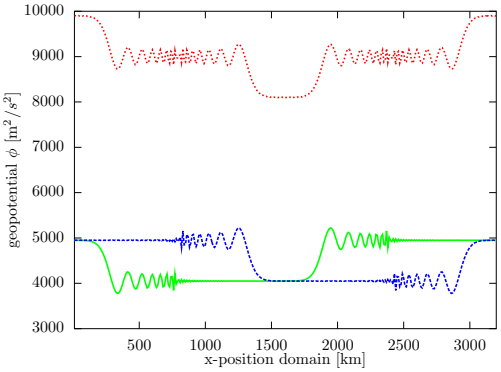
(b)



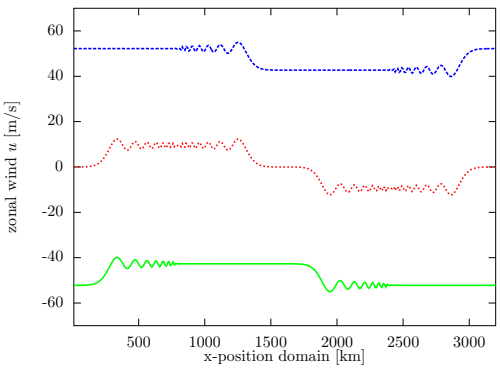
(c)



(d)

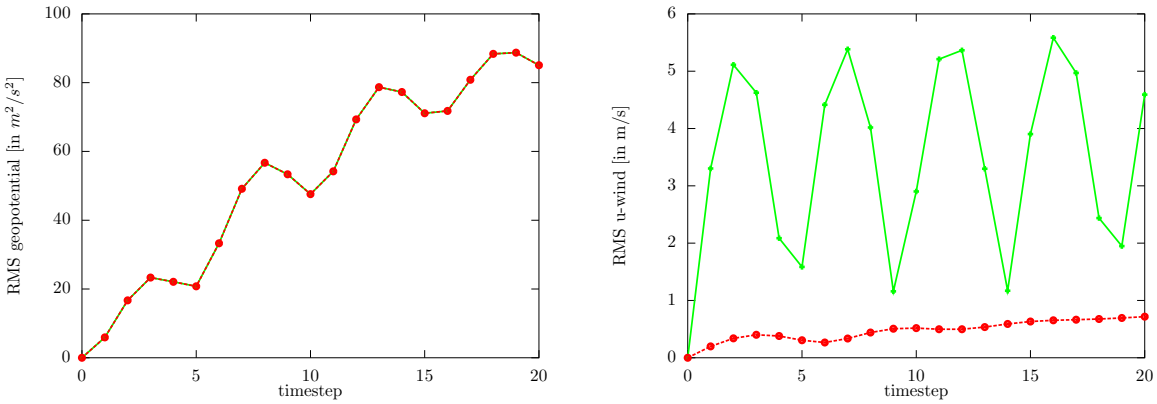


(e)



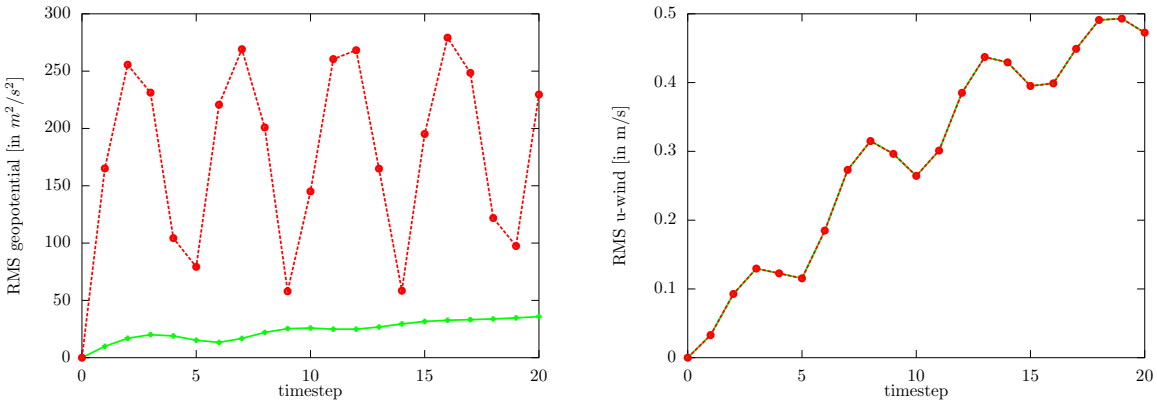
(f)

Figure 4.9: Identical to Figure 4.8 but now after 20 time steps.



(a) (b)

Figure 4.10: RMS error for the geopotential (a) and the wind velocity (b) for the geostrophic adjustment experiment. The used schemes are symmetric second-order FD SISL Z-grid schemes with DZ (green) and PO (red) reconstruction.



(a) (b)

Figure 4.11: Identical to Figure 4.10 but with an initial state consisting of a flat geopotential field and a discontinuity in the wind.

this can be seen in Figure 4.8 (d) where short waves appear in the u-field decomposition. PO-reconstruction results in u-components that are smaller than the analytical ones. This can be seen in Figure 4.8 (f) where the modi have a slope less steep than in Figure 4.8 (b). At the initial time the short scale waves of the DZ-reconstruction scheme are exactly canceled out by summing the two components. However, during time stepping both modi will propagate in opposite direction and the short scale u-fields will no longer cancel out. As a result, a noisy wind field that dominates the analytical wind field pops up in the Z-grid scheme, as illustrated in Figure 4.9 (d). The Z-grid scheme based on the PO reconstruction method will not give rise to u-noise.
Plots of the root mean square (RMS) difference of the symmetrical Z-grid schemes with the spectral (and thus analytical) solution are shown in Figure 4.10. For the geopotential field there is, indeed, no eigenvector problem. Both DZ and PO RMS errors are identical and growing slowly due to the difference in phase velocities between the Z-grid and spectral approach. However, for the wind velocity the situation is different. A large discrepancy between the DZ-based Z-grid scheme and the spectral scheme appears from the first time step on. This is explained by the overly presence of short scales in the wind decomposition that do not longer cancel out as soon as both waves start propagating in opposite direction. This also explains the periodic behavior in the RMS error. For the PO-based Z-grid scheme the RMS error for the wind is determined by the phase speed misrepresentation.
Based on the previous, one could erroneously conclude that the PO-reconstruction scheme is superior to the DZ-reconstruction. However, Figure 4.7 suggests that an initial state with a flat geopotential but a step in the u-velocity would lead to similar problems in the geopotential field of the PO-based Z-grid scheme. Figure 4.11 shows the RMS error of such a test and confirms the previous reasoning. The eigenvector-problem fingerprint is clearly visible in the RMS error for the geopotential field if PO is used.

For the 2D SWE with Coriolis force, there will be three eigenmodes: two IGW modi and one geostrophic mode. For the IGWs the previous effect is still present, but now the geostrophic eigenvector will also differ depending on the method used. This explains why in Figure 4.5 the stationary geostrophic field in the center is not identical for all schemes.

4.5 Nonlinear SWE tests

The previous sections started from the linearized version of the SWE to study dispersion properties and eigenmodes,. However, atmospheric modeling is a strongly non-

linear problem. In the remaining part of this chapter, we will therefore present results of numerical tests including advection and orography, two important nonlinear factors in NWP. We will investigate the following questions:

- Do we still see the fingerprint of the dispersion properties and eigenvector decomposition of the schemes if nonlinear terms are included in the numerical tests?
- The use of a local method permits the introduction of horizontally inhomogeneous terms (e.g., orography, reference temperature,...) in the SI part of the computations. Could this be an advantage for local schemes?

The tests will be undertaken in a 1D SWE context (what we want to investigate is not related to the dimension of the problem), but with the Coriolis force $f \neq 0$ and thus meridional velocity v included:

$$\frac{du}{dt} = -\frac{\partial \phi}{\partial x} + fv, \tag{4.65}$$

$$\frac{dv}{dt} = -fu, \text{ and} \tag{4.66}$$

$$\frac{d\left(\phi - \phi_{oro}\right)}{dt} = -\left(\phi - \phi_{oro}\right)\frac{\partial u}{\partial x}. \tag{4.67}$$

The 2TL SI Z-grid discretization looks like:

$$\begin{bmatrix} X - \Phi\left(\frac{\Delta t}{2}\right)^2 \frac{\partial^2}{\partial x^2} & 0 & 0 \\ \frac{f\Delta t}{2} & 1 & 0 \\ \frac{\Phi \Delta t}{2} & 0 & 1 \end{bmatrix} \begin{bmatrix} D^+ \\ Z^+ \\ \phi^+ \end{bmatrix} = \begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\frac{\partial^2}{\partial x^2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial x} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$\times \left(\begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\frac{\partial}{\partial x} \\ -\frac{f\Delta t}{2} & 1 & 0 \\ -\frac{\Phi \Delta t}{2}\frac{\partial}{\partial x} & 0 & 1 \end{bmatrix} \begin{bmatrix} u^0 \\ v^0 \\ \phi^0 \end{bmatrix} + \begin{bmatrix} R_u^{NL} \\ R_v^{NL} \\ R_\phi^{NL} \end{bmatrix} \right) \tag{4.68}$$

$$\begin{bmatrix} D^+ \\ Z^+ \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial x} \end{bmatrix} \begin{bmatrix} u^+ \\ v^+ \end{bmatrix}, \tag{4.69}$$

with $X = 1 + \left(\frac{f\Delta t}{2}\right)^2$, Φ a constant, homogeneous reference geopotential, and R_u^{NL}, R_v^{NL}, and R_ϕ^{NL} the nonlinear terms. If the explicit calculations are evaluated in the departure points, these nonlinear terms do not contain advection terms. On the other hand, if a Eulerian approach is followed, advection will be included in the nonlinear terms. The first line of Eq. (4.68) represents a Helmholtz equation to update divergence, the second and third line are used to calculate vorticity and geopotential. Equation

(4.69) defines how to retrieve the wind.
We will test four different discretization approaches:

1. **scheme 1 = spectral scheme**

$$\begin{bmatrix} X-\Phi\left(\frac{\Delta t}{2}\right)^2\mathcal{P}_{xx}^{sp} & 0 & 0 \\ \frac{f\Delta t}{2} & 1 & 0 \\ \frac{\Phi\Delta t}{2} & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} D^+ \\ Z^+ \\ \phi^+ \end{bmatrix} = \begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\mathcal{P}_{xx}^{sp} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathcal{P}_x^{sp} & 0 & 0 \\ 0 & \mathcal{P}_x^{sp} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\times \left(\begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\mathcal{P}_x^{sp} \\ -\frac{f\Delta t}{2} & 1 & 0 \\ -\frac{\Phi\Delta t}{2}\mathcal{P}_x^{sp} & 0 & 1 \end{bmatrix} \begin{bmatrix} u^0 \\ v^0 \\ \phi^0 \end{bmatrix} + \begin{bmatrix} R_u^{NL} \\ R_v^{NL} \\ R_\phi^{NL} \end{bmatrix} \right)$$

$$\begin{bmatrix} D^+ \\ Z^+ \end{bmatrix} = \begin{bmatrix} \mathcal{P}_x^{sp} & 0 \\ 0 & \mathcal{P}_x^{sp} \end{bmatrix} \begin{bmatrix} u^+ \\ v^+ \end{bmatrix}.$$

   This scheme represents the spectral dynamical core of the ALADIN model and results in excellent dispersion properties.

2. **scheme 2 = FD asymmetrical Z-grid scheme**

$$\begin{bmatrix} X-\Phi\left(\frac{\Delta t}{2}\right)^2\mathcal{P}_{xx}^{fd} & 0 & 0 \\ \frac{f\Delta t}{2} & 1 & 0 \\ \frac{\Phi\Delta t}{2} & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} D^+ \\ Z^+ \\ \phi^+ \end{bmatrix} = \begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\mathcal{P}_{xx}^{fd} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathcal{P}_x^{fd} & 0 & 0 \\ 0 & \mathcal{P}_x^{fd} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\times \left(\begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\mathcal{P}_x^{fd} \\ -\frac{f\Delta t}{2} & 1 & 0 \\ -\frac{\Phi\Delta t}{2}\mathcal{P}_x^{fd} & 0 & 1 \end{bmatrix} \begin{bmatrix} u^0 \\ v^0 \\ \phi^0 \end{bmatrix} + \begin{bmatrix} R_u^{NL} \\ R_v^{NL} \\ R_\phi^{NL} \end{bmatrix} \right)$$

$$\begin{bmatrix} D^+ \\ Z^+ \end{bmatrix} = \begin{bmatrix} \mathcal{P}_x^{fd} & 0 \\ 0 & \mathcal{P}_x^{fd} \end{bmatrix} \begin{bmatrix} u^+ \\ v^+ \end{bmatrix}.$$

   Replacing the spectral discretization in scheme 1 naively by a FD horizontal spatial discretization results in the scheme above. The scheme contains an asymmetry in the second-order derivative of the geopotential ϕ and, therefore, distorts the IGW eigenmodes. The wind reconstruction is based on the so-called DZ-method.

3. **scheme 3 = FD symmetrical Z-grid scheme**

$$\begin{bmatrix} X-\Phi\left(\frac{\Delta t}{2}\right)^2\mathcal{P}_{xx}^{fd} & 0 & 0 \\ \frac{f\Delta t}{2} & 1 & 0 \\ \frac{\Phi\Delta t}{2} & 0 & 1 \end{bmatrix}\begin{bmatrix} D^+ \\ Z^+ \\ \phi^+ \end{bmatrix} = \begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\mathcal{P}_{xx}^{fd} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \mathcal{P}_x^{fd} & 0 & 0 \\ 0 & \mathcal{P}_x^{fd} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\times \left(\begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\left(\mathcal{P}_x^{fd}\right)^{-1}\mathcal{P}_{xx}^{fd} \\ -\frac{f\Delta t}{2} & 1 & 0 \\ -\frac{\Phi\Delta t}{2}\mathcal{P}_x^{fd} & 0 & 1 \end{bmatrix}\begin{bmatrix} u^0 \\ v^0 \\ \phi^0 \end{bmatrix} + \begin{bmatrix} R_u^{NL} \\ R_v^{NL} \\ R_\phi^{NL} \end{bmatrix}\right)$$

$$\begin{bmatrix} D^+ \\ Z^+ \end{bmatrix} = \begin{bmatrix} \mathcal{P}_x^{fd} & 0 \\ 0 & \mathcal{P}_x^{fd} \end{bmatrix}\begin{bmatrix} u^+ \\ v^+ \end{bmatrix}.$$

This scheme restores symmetry by replacing the derivative operation on the geopotential $\mathcal{P}_x^{fd}\phi^0$ by $\left(\mathcal{P}_x^{fd}\right)^{-1}\mathcal{P}_{xx}^{fd}\phi^0$. This trick adds some complexity to the algorithm, but the eigenvalues now yield appropriate IGW dispersion. However, its eigenmodes differ from the analytical ones, as explained in section 4.4.

4. **scheme 4 = FD symmetrical A-grid scheme**

$$\begin{bmatrix} X-\Phi\left(\frac{\Delta t}{2}\right)^2\mathcal{P}_x^{fd}\mathcal{P}_x^{fd} & 0 & 0 \\ \frac{f\Delta t}{2} & 1 & 0 \\ \frac{\Phi\Delta t}{2} & 0 & 1 \end{bmatrix}\begin{bmatrix} D^+ \\ Z^+ \\ \phi^+ \end{bmatrix} = \begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\mathcal{P}_x^{fd}\mathcal{P}_x^{fd} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} \mathcal{P}_x^{fd} & 0 & 0 \\ 0 & \mathcal{P}_x^{fd} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\times \left(\begin{bmatrix} 1 & \frac{f\Delta t}{2} & -\frac{\Delta t}{2}\mathcal{P}_x^{fd} \\ -\frac{f\Delta t}{2} & 1 & 0 \\ -\frac{\Phi\Delta t}{2}\mathcal{P}_x^{fd} & 0 & 1 \end{bmatrix}\begin{bmatrix} u^0 \\ v^0 \\ \phi^0 \end{bmatrix} + \begin{bmatrix} R_u^{NL} \\ R_v^{NL} \\ R_\phi^{NL} \end{bmatrix}\right)$$

$$\begin{bmatrix} D^+ \\ Z^+ \end{bmatrix} = \begin{bmatrix} \mathcal{P}_x^{fd} & 0 \\ 0 & \mathcal{P}_x^{fd} \end{bmatrix}\begin{bmatrix} u^+ \\ v^+ \end{bmatrix}.$$

This fourth scheme is a symmetric scheme where the implicit second-order derivative $\mathcal{P}_{xx}^{fd}$ is evaluated as twice a first-order derivative $\mathcal{P}_x^{fd}\mathcal{P}_x^{fd}$. In this way a symmetric scheme is obtained such that there will be nor damping neither amplification. However, it can be verified that this scheme, despite being Z-grid in the sense that there is a transformation to divergence and vorticity, results in an A-grid type of IGW dispersion (negative group velocities for the shortest waves). Scheme 4 will not suffer from eigenvector distortions.

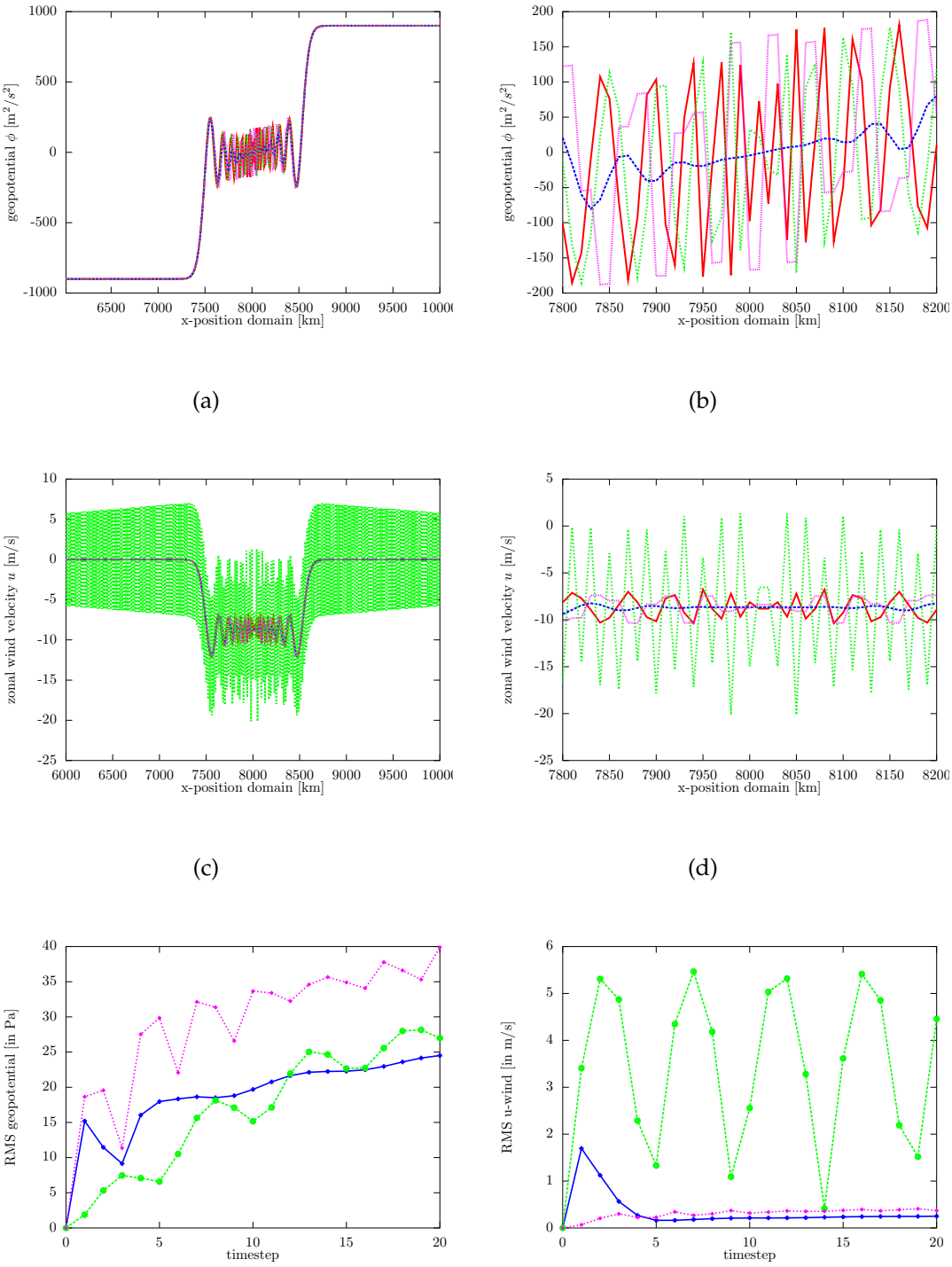


Figure 4.12: Plot of the geopotential and zonal wind after 20 time steps in a geostrophic adjustment test for the linear SWE. The parameters of the test are identical to the ones used in Figure 4.4 and Figure 4.6. Different schemes are used: scheme 1 (red), scheme 2 (blue), scheme 3 (green), and scheme 4 (magenta). Subfigures (e) and (f) show the evolution of the RMS error with respect to scheme 1.

As a reference point for the following experiments, we will first repeat the linear geostrophic adjustment test, used in Figure 4.2 and Figure 4.6, for the 4 schemes. Running the schemes with $R_u^{NL} = R_v^{NL} = R_\phi^{NL} = 0$ and in Eulerian mode with a geopotential step as initial state results in Figure 4.12. We remark:

- that there is no qualitative difference between using a linear FE scheme (Figure 4.4 and 4.6) and a second-order FD scheme (Figure 4.12). The waves in the DZ-based symmetrical scheme combined with linear FE (green in Figure 4.4 and 4.6) and second-order FD (green in Figure 4.12) are faster and slower, respectively, than the ones of the spectral schemes. This agrees with the dispersion plot of Figure 4.2.
- that scheme 4 results indeed in an A-grid type of dispersion. The ripples on the long waves are a token of the short scale noise propagating at large speeds in the opposite direction. This behavior appears only for scheme 4 and agrees with the A-grid dispersion analysis of Figure 3.4.
- that scheme 3 suffers from the eigenvector problem for the wind field (Figure 4.12 (c) and (d)), as can be seen by the abrupt rise of the RMS error for u from the first time step on (Figure 4.12 (f)). The linear character of this test makes that the problem cannot spill over on other fields.
- that the asymmetrical Z-grid scheme 2 does not have this noise, which suggests that for this scheme the damping is dominant. It was verified that during the first time steps the short scale effects due to the eigenvectors are present for scheme 2. This can be seen in the RMS error for u (Figure 4.12 (f)), which is showing a maximum during the first time steps for scheme 2. However, after more or less 5 time steps all short scale waves are damped (no periodic variation in the RMS error compared with scheme 2 and 3). One may think that in terms of RMS error the unphysical damping is advantageous, because it removes the shortest scales that are otherwise represented wrongly. However, one should avoid compensating errors.
- that both scheme 3 and scheme 4 result in short scale waves that are standing still in the middle of the domain. This agrees with the fact that for both schemes a certain part of the spectrum has a group velocity near 0; for scheme 3 this happens around wavelengths $\lambda \approx 2\Delta x$ and for scheme 4 around $\lambda \approx 4\Delta x$. Due to the damping, this is not found back in scheme 2.

4.5.1 Nonlinear test without orography

The nonlinear terms of the SWE are introduced in the discretized equation set (4.68) via R_u^{NL}, R_v^{NL}, and R_ϕ^{NL}. Depending on whether SL advection is used, they will or will not include advection terms. The nonlinear terms are identical for all four schemes and the schemes only differ in the discretization of the linear part of the equations.
Different methods exist to include nonlinear residuals. In what follows the iteratively centered implicit (ICI) approach, which was advocated for in the nonhydrostatic ALADIN documentation of Bénard and Masek [7], is opted for. With this approach the system of equations is solved multiple times. In analogy with Eq. (2.3) a 2TL ICI SISL method with N iterations can be summarized as

$$\begin{aligned}
(\mathcal{I} - \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{F}_A^{+(0)} &= (\mathcal{I} + \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{F}_{D^{(0)}}^0 + \frac{\Delta t}{2}(\mathcal{M} - \mathcal{L}^*)\left(\mathbf{F}_{D^{(0)}}^0\right) + \frac{\Delta t}{2}(\mathcal{M} - \mathcal{L}^*)\left(\mathbf{F}_A^0\right) \\
(\mathcal{I} - \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{F}_A^{+(i)} &= (\mathcal{I} + \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{F}_{D^{(i)}}^0 + \frac{\Delta t}{2}(\mathcal{M} - \mathcal{L}^*)\left(\mathbf{F}_{D^{(i)}}^0\right) \\
&+ \frac{\Delta t}{2}(\mathcal{M} - \mathcal{L}^*)\left(\mathbf{F}_A^{+(i-1)}\right), \quad \forall i = 1, .., N
\end{aligned}$$

where i is the iteration index and $\mathbf{F}_A^{+(N)}$ the updated field. For the sake of simplicity, the physics tendencies are not taken into account. Different methods exist to determine the location of the departure points. For our purpose the following direct method was used:

$$\begin{aligned}
x_D^{(0)} &= x_A - u_A^0 \Delta t \\
x_D^{(i)} &= x_A - \left(u_A^0 + u_A^{+(i-1)}\right)\frac{\Delta t}{2}, \quad \forall i = 1, .., N.
\end{aligned}$$

Cubic interpolation is used to evaluate functions in the departure points.
In case advection is evaluated in an Eulerian way, the 2TL ICI SI method looks like:

$$\begin{aligned}
(\mathcal{I} - \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{F}^{+(0)} &= (\mathcal{I} + \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{F}^0 + \Delta t(\mathcal{M} - \mathcal{L}^*)\left(\mathbf{F}^0\right) \\
(\mathcal{I} - \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{F}^{+(i)} &= (\mathcal{I} + \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{F}^0 + \frac{\Delta t}{2}(\mathcal{M} - \mathcal{L}^*)\left(\mathbf{F}^0\right) \\
&+ \frac{\Delta t}{2}(\mathcal{M} - \mathcal{L}^*)\left(\mathbf{F}^{+(i-1)}\right), \quad \forall i = 1, .., N
\end{aligned}$$

with the advection terms included in the nonlinear residuals $(\mathcal{M} - \mathcal{L}^*)$.

Based on Eq. (4.65)-(4.67), one can write out the expressions for the nonlinear terms.

For the Eulerian variants of the different schemes this yields:

$$\begin{aligned}
R_u^{NL(i)} &= -\frac{\Delta t}{2}\left[\left(u\frac{\partial u}{\partial x}\right)^0 + \left(u\frac{\partial u}{\partial x}\right)^{+(i-1)}\right], \\
R_v^{NL(i)} &= -\frac{\Delta t}{2}\left[\left(u\frac{\partial v}{\partial x}\right)^0 + \left(u\frac{\partial v}{\partial x}\right)^{+(i-1)}\right], \text{ and} \\
R_\phi^{NL(i)} &= -\frac{\Delta t}{2}\left[\left(u\frac{\partial\left(\phi-\phi_{oro}\right)}{\partial x}\right)^0 + \left(u\frac{\partial\left(\phi-\phi_{oro}\right)}{\partial x}\right)^{+(i-1)}\right] \\
&- \frac{\Delta t}{2}\left[\left(\left(\phi-\phi_{oro}\right)D\right)^0 + \left(\left(\phi-\phi_{oro}\right)D\right)^{+(i-1)}\right], \quad \forall i = 1,..,N
\end{aligned} \tag{4.70}$$

where a reference state in rest is assumed (otherwise one should replace u by the total wind $u+U$). For $i=0$ the previous expression is replaced by:

$$\begin{aligned}
R_u^{NL(0)} &= -\Delta t\left(u\frac{\partial u}{\partial x}\right)^0, \\
R_v^{NL(0)} &= -\Delta t\left(u\frac{\partial v}{\partial x}\right)^0, \text{ and} \\
R_\phi^{NL(0)} &= -\Delta t\left(u\frac{\partial\left(\phi-\phi_{oro}\right)}{\partial x}\right)^0 - \Delta t\left(\left(\phi-\phi_{oro}\right)D\right)^0.
\end{aligned}$$

In case the SL approach is used, one obtains

$$\begin{aligned}
R_u^{NL(i)} &= R_v^{NL(i)} = 0 \text{ and} \\
R_\phi^{NL(i)} &= -\frac{\Delta t}{2}\left[\left(\left(\phi-\phi_{oro}\right)D\right)^0_{D^{(i)}} + \left(\left(\phi-\phi_{oro}\right)D\right)^{+(i-1)}_A\right] \\
&+ \left(\phi_{oro}\right)_A - \left(\phi_{oro}\right)_{D^{(i)}}, \quad \forall i = 1,..,N
\end{aligned} \tag{4.71}$$

and for $i=0$

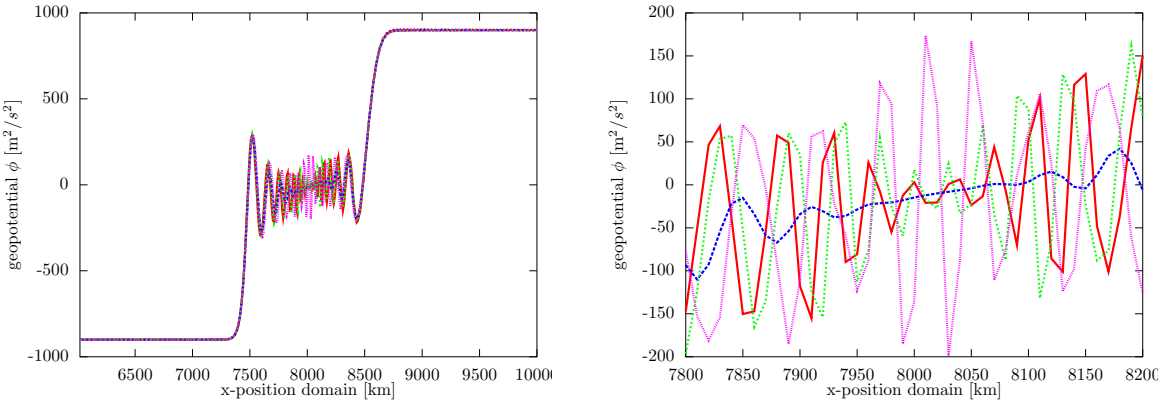
$$\begin{aligned}
R_u^{NL(0)} &= R_v^{NL(0)} = 0 \text{ and} \\
R_\phi^{NL(0)} &= -\frac{\Delta t}{2}\left[\left(\left(\phi-\phi_{oro}\right)D\right)^0_{D^{(0)}} + \left(\left(\phi-\phi_{oro}\right)D\right)^0_A\right] + \left(\phi_{oro}\right)_A - \left(\phi_{oro}\right)_{D^{(0)}}.
\end{aligned}$$

Let us repeat the geostrophic adjustment test but this time with the inclusion of non-linear terms.
With Eulerian advection the adjustment test results in Figure 4.13. The fields are truncated at two-thirds of the spectrum $k_{trunc} = \frac{2}{3}k_{max}$ to prevent instabilities caused by a build-up of small scale noise by aliasing. This is called quadratic truncation because it avoids aliasing due to quadratic terms. Quadratic truncation is not sufficient to remove the stationary $4\Delta x$-wave of scheme 4. However, the eigenvector problem that

was remarked before if scheme 3 was used is controlled in this test. We thus get, based on the RMS scores, the following order for the quality of the forecast: scheme 3, scheme 2 and scheme 4.
If, instead of the Eulerian strategy, advection is treated in a SL way, the adjustment test results in Figure 4.14 (with quadratic truncation) and Figure 4.15 (without truncation). Both experiments used only one iteration of the ICI scheme. The interpolation inherent to a SL scheme adds some diffusion, and this can be seen in the smooth evolution of the RMS error for all schemes. In principle, truncation is no longer necessary in a SL approach. Advection is, indeed, the dominant source for the aliasing problem. It can be remarked that without truncation the eigenvector problem is again present. This time the corresponding RMS error in scheme 3 decreases after some integrations (due to the inherent diffusion), but it still spoils the scheme and makes it less good than scheme 4. The dispersion issues that were noted earlier in the linear tests are mostly gone in the nonlinear tests. It turns out that quadratic truncation eliminates the part of the spectrum that is most vulnerable for these issues. The $4\Delta x$ stationary waves of scheme 4 are the exception as they are still visible for both Eulerian and SL tests. The dampening of the IGWs, due to the asymmetry of scheme 2, remains in the nonlinear tests. Furthermore, comparing Figures 4.13 and 4.14 illustrates the smoothing typical for the interpolations used in the SL schemes.

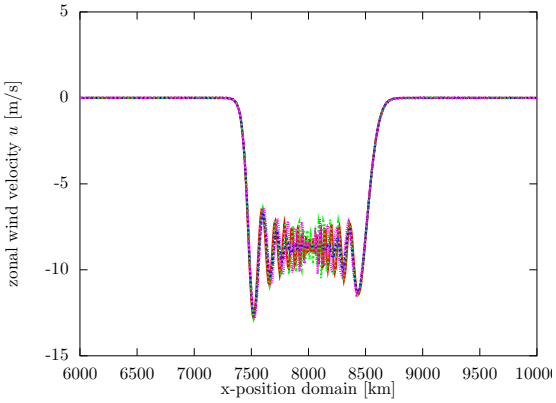
4.5.2 Nonlinear test with orography

Including orography adds some extra complexity to the nonlinear right hand sides. Before testing the different schemes over orography, one should realize that an alternative treatment of these orography terms is possible, at least, if a local method (FD, FE,...) is used for the spatial discretization. The orography field $\Phi_{oro}(x)$ is known and it is possible to include terms where orography is multiplied with a prognostic field in the implicit part of the time discretization. In case of the SWE, one could treat the $\Phi_{oro}D$ term in an implicit way. Doing so will result in a Helmholtz solver containing the spatially varying orography. This approach is advantageous because it reduces the magnitude of the nonlinear terms, which are a potential source for instabilities. The implicit treatment of orography terms is no option for spectral algorithms because the inhomogeneous orography would result in a Helmholtz problem with full matrices in spectral space. This would make the spectral method highly inefficient. In Appendix B a local SI Z-grid scheme that treats the term $\Phi_{oro}(x)\,D$ implicitly is presented.
A good academic test consists of simulating the 1D flow over a Gaussian hill $\Phi_{oro}(x)$. As initial state we assume a $5\,\mathrm{m\,s^{-1}}$ zonal wind and an orography geopotential about half the height of the reference geopotential $\Phi = 9000\,\mathrm{m^2\,s^{-2}}$, as shown in Figure 4.16 (a) and (b). For the experiment a time step $\Delta t = 150\,\mathrm{s}$ and a grid distance $\Delta x = 10\,\mathrm{km}$

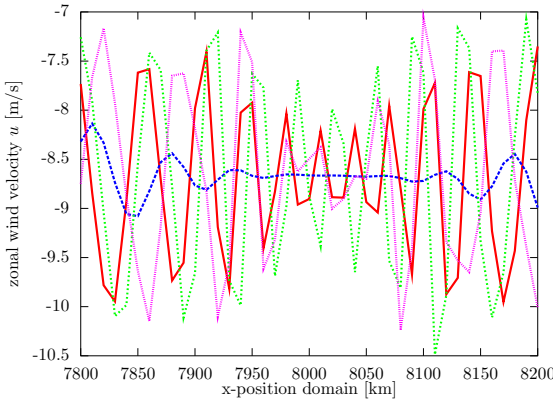


(a)

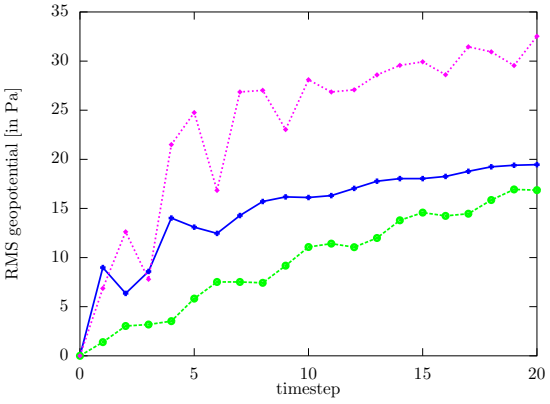
(b)



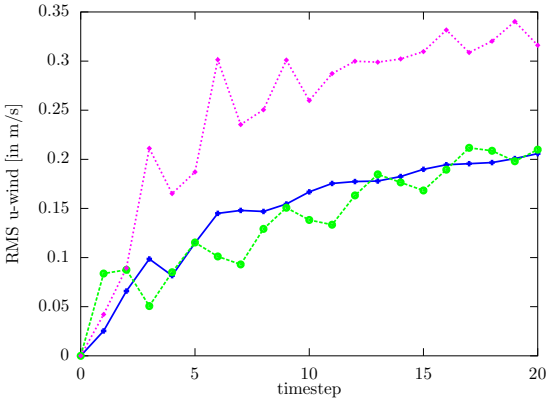
(c)



(d)



(e)



(f)

Figure 4.13: Identical to Figure 4.12 but with nonlinear terms included. Advection is treated in a Eulerian way and 1 iteration was used for the ICI scheme. The fields are truncated at $k_{trunc} = \frac{2}{3}k_{max}$.

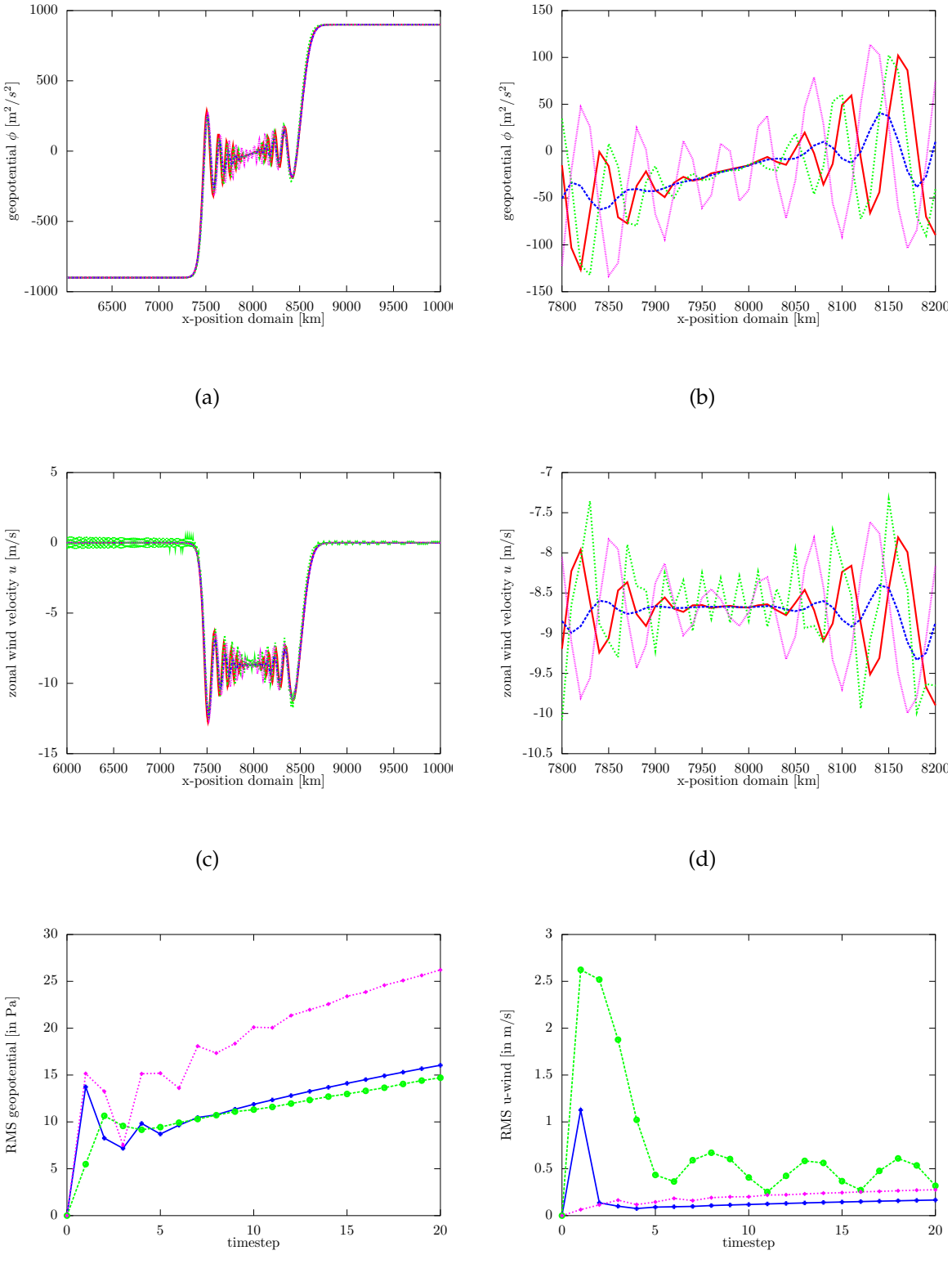


Figure 4.14: Identical to Figure 4.12 but with nonlinear terms included. Advection is treated in a SL way (cubic interpolation) and 1 iteration was used for the ICI scheme. The fields are not truncated.

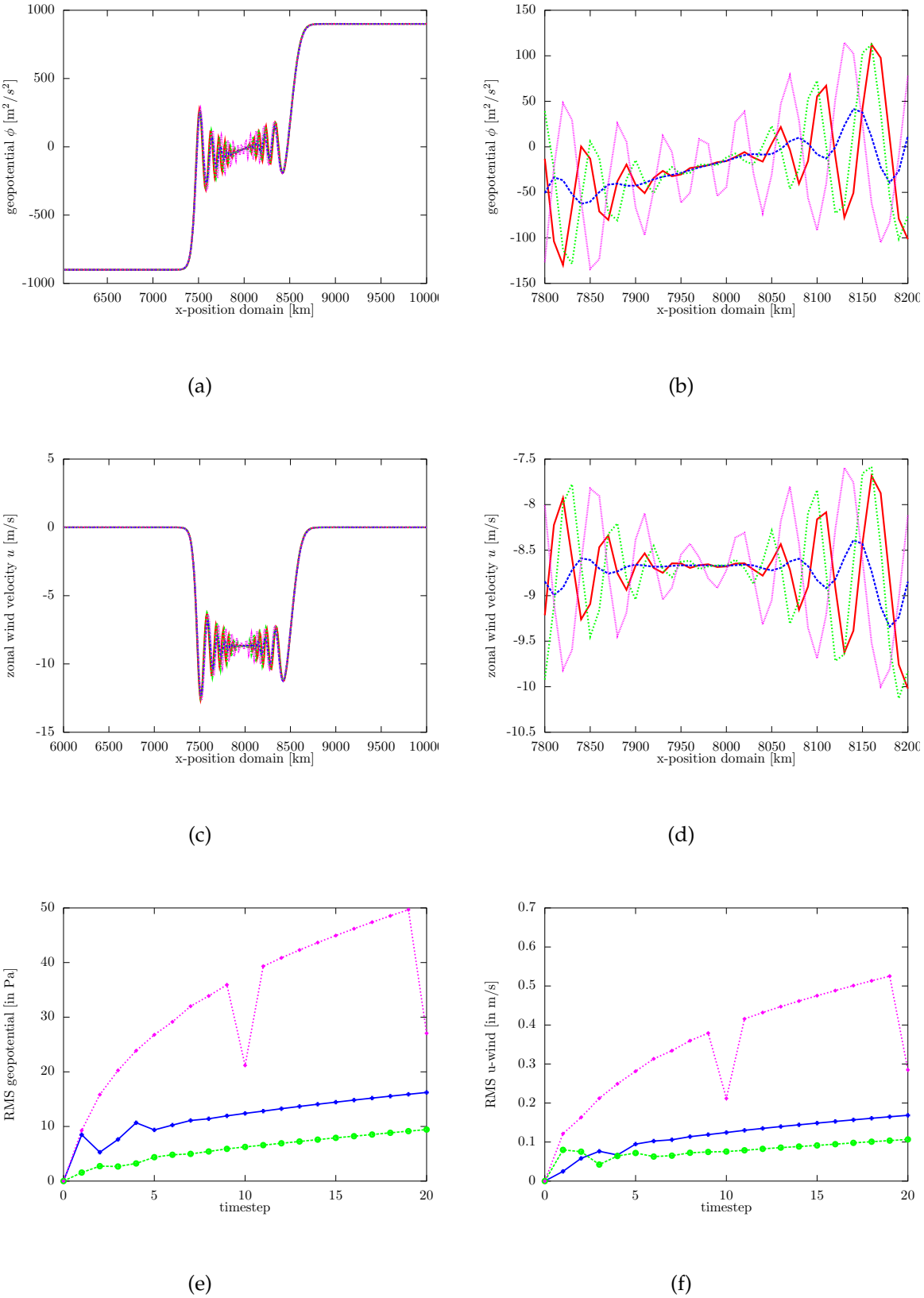


Figure 4.15: Identical to Figure 4.12 but with nonlinear terms included. Advection is treated in a SL way (cubic interpolation) and 1 iteration was used for the ICI scheme. The fields are truncated at $k_{trunc} = \frac{2}{3}k_{max}$.

are used. Advection is treated in a Eulerian way and quadratic truncation is assumed. Tests were done for the standard version of scheme 3 and for scheme 3 but with $\Phi_{oro} D$ treated implicitly. The horizontal spatial discretization is based twice on a second-order FD method. Figure 4.16 (c) and (d) show the results after 100 time steps if 10 ICI iterations are used. These plots confirm that a sufficient number of iterations makes that the two approaches converge to each other and becomes indistinguishable. In practice, one tries to limit the number of iterations in order to minimize the computation and energy costs. It is interesting to study the convergence rate of the two approaches to this 10 iteration solution. Figure 4.16 (e) and (f) show the difference between the results after 0, 1, 2, and 10 iterations for both approaches. Despite being slightly better, we see that without iteration both the implicit (green) and explicit (red) orography scheme are quite far from the converged solution. After 1 iteration the differences are strongly reduced for both approaches and 2 iterations are sufficient to bring all differences down to nearly 0. Whereas the blue curve of the explicit scheme with 1 iteration is still distinguishable, this is not the case if the implicit orography approach is used with 1 iteration. Adding the inhomogeneous $\Phi_{oro}(x)$ to the implicit calculations gives slightly faster convergence. The differences are not so large, and this is probably due to the fact that the $\Phi_{oro}(x) D$ is not the dominant nonlinear term. We conclude that both approaches work, but that the errors for the same number of iterations are slightly lower if the nonlinear residuals are reduced by the implicit orography treatment.
This test was repeated in the case advection is accounted for by a SL approach. A cubic reconstruction formula is used for the interpolations to the departure points. Results are shown in Figure 4.17, again the implicit orography treatment converges faster. This time the errors of the implicit scheme without iteration are surprisingly low: this is because orography is already accounted for (because it is implicit) and advection is better approximated by a SL treatment than in the Eulerian way if no iteration is used.

4.6 Conclusion

After the dispersion study of Chapter 3, the Z-grid approach came out as a good way to introduce local discretization schemes within the current spectral ALADIN approach. It shows excellent IGW properties and fits well in the ALADIN time step organization, which is also based on a divergence-vorticity approach. In this chapter a closer investigation of SISL Z-grid schemes is undertaken.
Two nuances related to the use of Z-grid schemes come out of the analysis in this chapter:

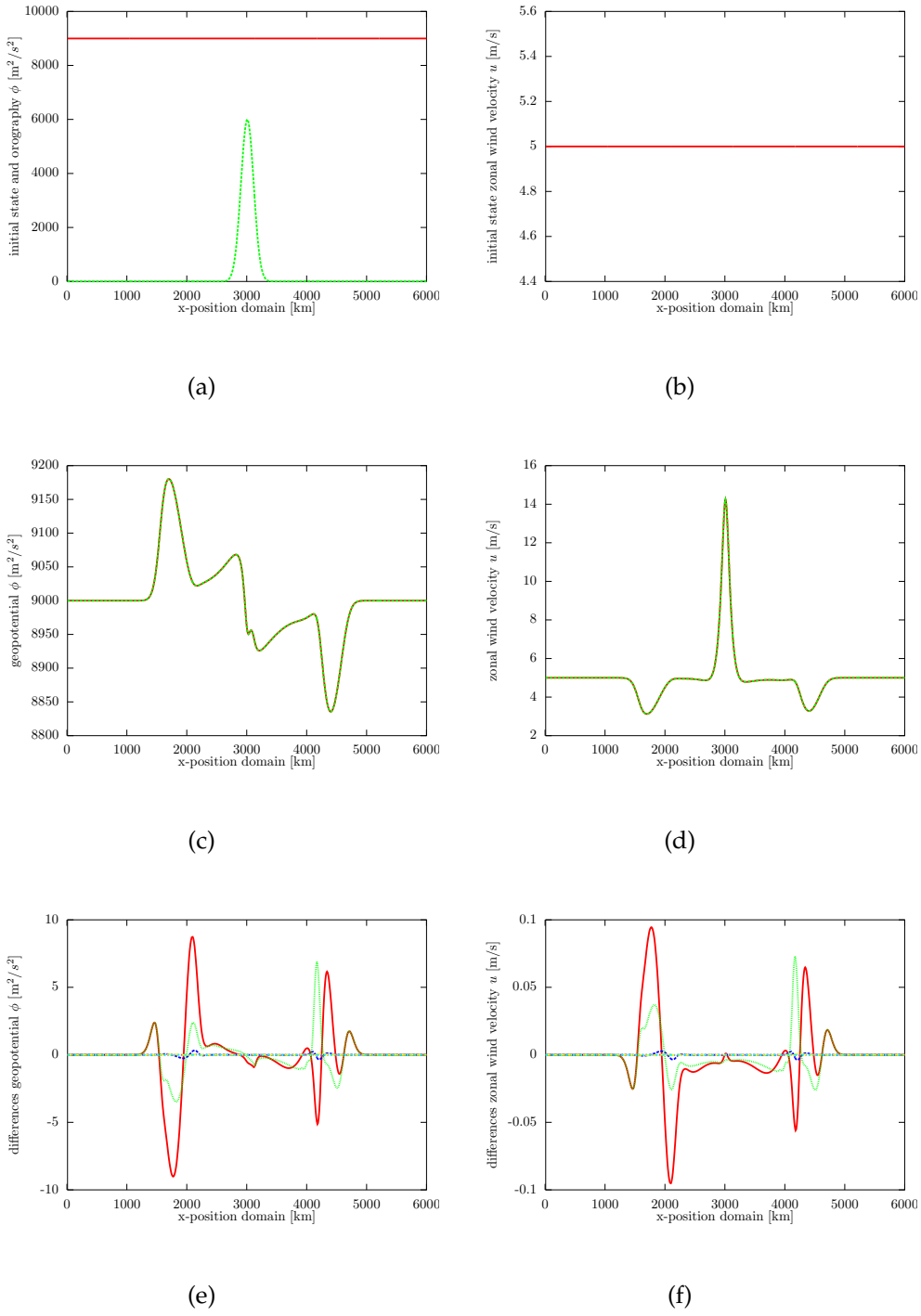


Figure 4.16: Different implementations of the symmetric Z-grid scheme (scheme 3) are used to simulate the flow over a hill. The initial state of the geopotential and zonal wind is shown in (a) and (b) together with the orography $\Phi_{oro}(x)$. The parameters used for the simulation are: $\Delta t = 150\,\text{s}$, $\Delta x = 10\,\text{km}$, and $\Phi = 9000\,\text{m}^2\,\text{s}^{-2}$. Figure (c) and (d) show the geopotential and zonal wind velocity after 100 time steps for scheme 3 with explicit (red) and implicit treatment of orography (green) if 10 ICI iterations are used. Figure (e) and (f) show the differences with respect to this converged solution of the schemes without iteration (red = expl oro, green = impl oro), with 1 iteration (blue = expl oro, yellow = impl oro) and with 2 iterations (magenta = expl oro, cyan = impl oro).

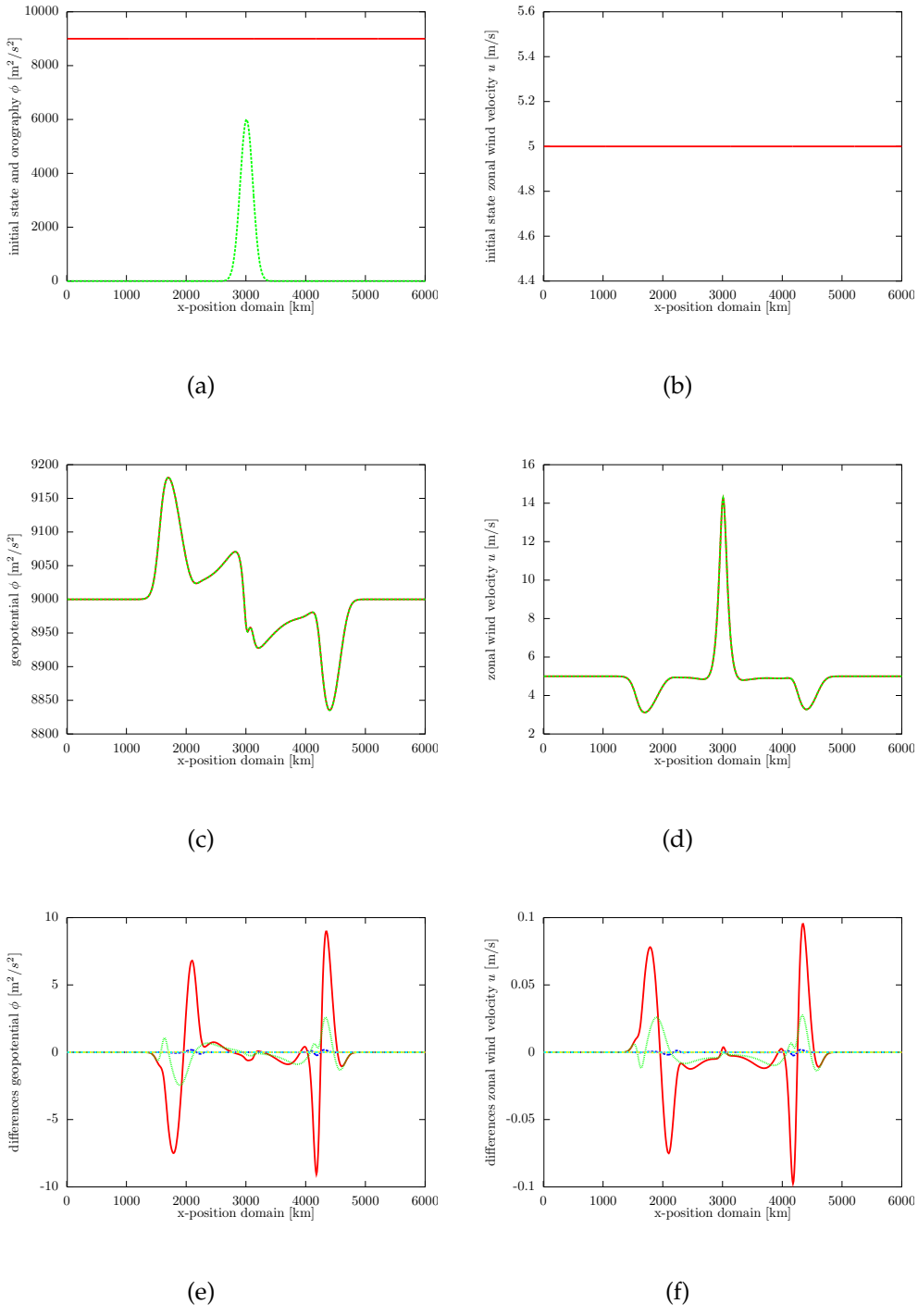


Figure 4.17: Similar as Figure 4.16 but this time with SL advection (cubic interpolation).

1. SI schemes that are asymmetric, in the sense that the spatial operations on a field are different at time t^+ and t^0, have unphysical dispersion relations. This is a general truth for SI schemes but in practice this situation is rarely encountered. However, within the Z-grid framework, where transformations between wind and vorticity and divergence are commonplace, one could unwarily introduce asymmetries as is the case for the schemes presented in ST86 and TS12. It is shown that in this case the appropriate dispersion properties make place for a dispersion relation with unphysical dampening of the shortest waves. To restore symmetry, care should be taken in the formulation of the different operators. By introducing more complex derivative operators, symmetric SI Z-grid schemes are constructed that do no longer suffer from this issue.
2. The eigenvectors of Z-grid schemes, which determine the decomposition of an initial state onto the different eigenmodes, differ from the analytical ones for the short scale part of the spectrum. In practice, this results in short scale noise in the velocity or geopotential field. This issue is inherent to local Z-grid schemes and does not appear in A-grid or spectral Z-grid schemes.

As soon as one leaves the linear tests and adds extra complexity to the test, for instance advection, the fingerprints of the previous issues become less visible. Including nonlinear terms strongly impacts the shortest scale part of the spectrum: a Eulerian treatment of advection terms necessitates truncating the shortest scale waves to prevent instabilities, whereas a SL method does need interpolations to departure points smoothing out the shortest scales. It is found that most of the problems clearly visible in the linear tests are no longer seen in the nonlinear tests. This raises questions about the importance of the short scale dynamics. The relevance of dispersion properties is maybe overestimated.

It turns out to be impossible to find a local, unstaggered scheme not suffering from any problem. Using the A-grid scheme results in negative group velocities for the IGWs whereas the symmetric Z-grid schemes suffer from a wrong decomposition on the IGW eigenmodes. However, implementing an A-grid approach is easier and may therefore be preferred over the Z-grid.

The high-order character of the spectral method makes it very hard, if not impossible, to construct a local method that beats the spectral approach in terms of accuracy. The strength of local methods resides in their flexibility compared to spectral schemes. It is, for example, possible to include inhomogeneous terms in the SI treatment and thus Helmholtz equation if a local discretization is used without complicating the solver too much. This was illustrated with a SWE test where the $\Phi_{oro}(x) D$ term was included in the SI solver resulting indeed in a faster convergence (in terms of the number of ICI iterations).

This analysis confirms that it is difficult to make one choice for the spatial discretization and, therefore, motivates the need for a modular approach. Having one SISL framework where both local A-grid, local Z-grid, and spectral schemes fit, would be very interesting. Next chapter will present such a framework that will be used to make a scientific comparison of these different spatial discretization approaches within the ALADIN model.

Local horizontal spatial discretization tests with the ALADIN model

5

Until now, the properties of local horizontal spatial discretization A-grid and Z-grid schemes were discussed based on theoretical analysis and toy model tests solving the SWE. In section 5.1 the implementation of a local solver in the ALADIN model is discussed. An overview of the changes that are necessary to include a local spatial discretization method within the earlier introduced time step organization is given. However, adding a new local solver to the existing ALADIN model is a very technical task exceeding the ambitions of this work.

Fortunately implementing a solver is not necessary to investigate the scientific impact of a local spatial discretization method in ALADIN. Section 5.2 explains how the current code provides a testbed to compare the spectral SISL approach with the FD A-grid SISL and FD asymmetrical Z-grid SISL schemes. We describe how, by modifying the spectral responses, the ALADIN model can be used to emulate FD strategies for the horizontal spatial discretization. We study the impact of the use of A-grid and Z-grid SISL schemes compared with the current spectral SISL strategy. The study is undertaken for the ALADIN model in adiabatic mode (section 5.3) and ALARO mode (section 5.4). Apart from conclusions about the relevance of eigenvalue and eigenvector properties, this set-up also gives the opportunity to study the impact of the order of accuracy. Finally, conclusions are presented.

5.1 Solving a local SISL scheme

5.1.1 Time step organization of a local SISL scheme

Introducing a local spatial discretization method in the ALADIN algorithmics, which was described in Table 2.2, will have the following consequences:

- Transformations between spectral and grid point space (steps 1 and 8 in Table 2.2) become redundant, at least if we switch to a FD spatial discretization. A FE scheme still needs transformations between coefficient and grid point space, but

these will not require costly global communications of the transpositions needed in the current spectral model[1].

- The current nonhydrostatic ALADIN model requires the solution of one diagonal Helmholtz problem for each vertical model level [7]. Using a local spatial discretization method will in general result in an asymmetric 3D Helmholtz problem[2]. A discussion about solving such a problem follows in 5.1.2.
- The calculation of derivatives (needed for example for the trajectory calculations) can no longer be done in spectral space, but should be done with the local method. The same holds for diffusion, currently represented by a diagonal operation in spectral space.

Table 5.1 presents an overview of the time step organization if a FD scheme would be used. Whether an A-grid or Z-grid approach is chosen, does not impact the workflow too much. Differences will be found mainly in step 7 of Table 5.1. The A and Z-grid variants differ in the discretized form of the Helmholtz operator $(\mathcal{I} - \frac{\Delta t}{2}\mathcal{L}^*)$ and the way to retrieve the updated wind velocities (via a Poisson equation from divergence-vorticity for Z-grid or directly via the momentum equations for A-grid). The LBC coupling (step 6 in Table 5.1) can still be done with the Davies method, but there is no longer a need to biperiodize the fields.
Exploiting the flexibility of the local discretization, by including inhomogeneous terms in the implicit treatment (e.g., a horizontally inhomogeneous reference temperature), will also change the explicit computations in step 4 of Table 5.1.

5.1.2 Local solvers for 3D Helmholtz problems

A local SISL discretization method will in general result in an asymmetric 3D Helmholtz problem for one of the prognostic variables. Remembering the notations introduced in 1.3.1, one gets

$$\mathcal{H}(x,y,z,t)\, f^{+} = R, \tag{5.1}$$

with $\mathcal{H}(x,y,z,t)$ the inhomogeneous 3D Helmholtz operator, f^{+} the updated prognostic variable, and R the known right hand side. Spatial discretization will translate this problem into an asymmetric matrix problem

$$\mathbf{H}\mathbf{f}^{+} = \mathbf{R}, \tag{5.2}$$

[1] There may be one exception: the calculation of some of the derivative operators of the symmetric Z-grid schemes can be quite expensive.

[2] It may be that keeping the terms included in the implicit operator identical to the current situation, could result even for local methods in 1 Helmholtz problem on each level. However, as soon as inhomogeneous terms will be included in the implicit part, one will end up with a 3D Helmholtz problems.

Table 5.1: The different steps of interest carried out during one time step integration of a FD SISL model. The equivalent of each term in Eq. (2.3) is given in the second column of the table (based on Table 2 in [117]).

	ALADIN time step organization	
1	calculate physics in arrival points	$\mathcal{P}(\mathbf{U}_A^0)$
2	update tendencies	
3	compute SL departure points D and do interpolations	
4	compute explicit part dynamics	$(\mathcal{I} + \frac{\Delta t}{2}\mathcal{L}^*)\mathbf{U}_D^0 + \Delta t(\mathcal{M} - \mathcal{L}^*)\left(\tilde{\mathbf{U}}\right)$
5	add all tendencies	$\mathbf{R}_{lam}$
6	lateral boundary coupling	$\mathbf{R}_{tot} = \alpha\mathbf{R}_{host} + (1-\alpha)\mathbf{R}_{lam}$
7	local solver to update fields	$\mathbf{U}_A^+ = (\mathcal{I} - \frac{\Delta t}{2}\mathcal{L}^*)^{-1}\mathbf{R}_{tot}$

where $\mathbf{H}$ has dimensions of the order 10^6 by 10^6 for a LAM. This huge matrix is fortunately sparse, meaning that it contains only a few nonzero elements for each line, as illustrated for a small 2D problem in Figure 5.1(b). The higher the order of accuracy, the more nonzero elements will be present in the matrix. One should tone down the apparent simplicity of the spectral representation (Figure 5.1(a)) as transformations to spectral space come not for free, as discussed in 2.3.2.

The need to solve systems like Eq. (5.2) efficiently on HPC systems is encountered in many branches of science. That is why many algorithms are developed to solve such problems. There exist direct sparse matrix solvers, but they are generally considered as not very suitable for large matrices on massively parallel machines. For dimensions like the ones met in NWP one should use iterative methods. Two main categories of iterative solvers can be distinguished: *Krylov* and *multigrid* solvers [69]. To limit the number of iterations needed, *preconditioning* is used to reformulate the problem in a matrix problem with a faster convergence. One should find a balance between the complexity of the preconditioner and the convergence speed-up of the Helmholtz equation [107]. There is an unlimited set of potential preconditioners and there does not exist a procedure to find the best preconditioner. A preconditioner can be derived from the physical equations underlying the linear problem or constructed directly from the matrix $\mathbf{H}$.

The UK Met Office has many years of experience on local SISL approaches. In their models, first New Dynamics [28] and now ENDGame [40], different inhomogeneous terms are treated in an implicit way resulting in an asymmetric 3D Helmholtz problem. Their New Dynamics solver was a Krylov one based on the methodology described in

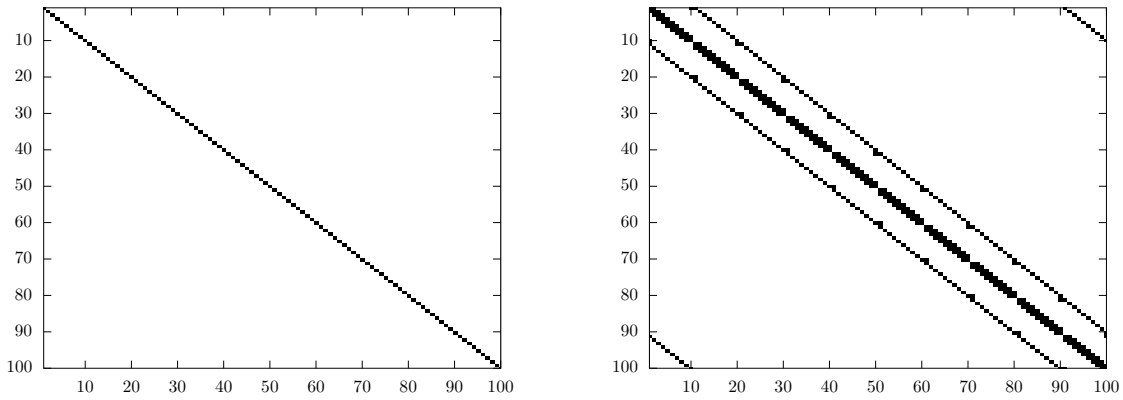


Figure 5.1: Nonzero elements of matrices representing a constant coefficient Helmholtz problem for a 10 by 10 domain in spectral space (a) and for a second-order FD discretization (b). Periodicity at the boundaries was assumed to create the matrices.

[99][3], but more recently they also investigated multigrid solvers for the ENDGame dynamics [69]. According to recent work of Muller, 3D Helmholtz solvers scale well on massively parallel machines [69].

Let us finish this section by coming back on the extra computations needed in a Z-grid method compared to the A-grid method. The need to retrieve the wind velocity components from divergence and vorticity results every time step in a Poisson problem. This problem does also exist in the current spectral SISL ALADIN dynamical core, but there its solution came for free (given that the transformations were needed anyway for the Helmholtz equation). If a local method is used, another inversion problem needs to be solved. The Poisson problem is a special case of the Helmholtz problem and similar solvers can be used. Good scalability can be obtained for Poisson problems, as illustrated recently in [77] that describes a multigrid solver scaling well on petascale computers. However, scalability of this problem seems less important because the wind retrieval is a 2D process, because there is no coupling between different levels. This automatically makes this problem an order of magnitude smaller than the 3D Helmholtz problem.

Development of sparse solvers is an active topic of research on its own, which falls outside the scope of this thesis. As will be demonstrated in the next section, it is nevertheless possible to evaluate the scientific impact of the use of local methods with the spectral ALADIN model.

[3] WRF was previously based on this approach

5.2.1 Methodology of the tests

As explained earlier, the choice of the horizontal spatial discretization cannot be separated from other decisions. A spectral spatial discretization, for example, is always combined with a SI time discretization in order to obtain an efficient scheme. It is therefore difficult to compare in a clean way different approaches for the horizontal spatial discretization. Despite the wide diversity in spatial discretization methods used in today's operational models, an intercomparison would not lead to conclusions about the spatial discretization as these models differ over a whole range of properties.
However, there are some good reasons to compare different spatial discretizations. In contrast to the extensive collection of papers published about the SWE dispersion properties of numerical schemes, there has to the author's knowledge never been a study about the relevance of these properties in a complete atmospheric model. Such a study would need a framework where one could choose between different spatial discretization schemes in an otherwise unchanged NWP model. In what follows, it will be explained that ALADIN offers such a testbed for spatial discretization schemes.

The set-up of the ALADIN model was discussed earlier. Figure 2.4 presents an overview of the general organization of the computations of the global IFS/ARPEGE model during one time step, while Table 2.2 outlines in more detail the time step organization. Implementing a local spatial discretization variant within the SISL ALADIN model would necessitate developing a new solver, as discussed in 5.1.2. Within ALADIN, where the spectral method assumes a decomposition into bi-Fourier functions, one could mimic a local method by simply replacing the spectral responses by the responses of a local method. This permits to emulate for instance a FD SISL model by intervening only in step 9 of Table 2.2. This simple but powerful approach permits to investigate how ALADIN forecasts look like if the spectral method is replaced by a local discretization scheme. Such a study can offer insight in the impact of the spatial discretization on the results of a model. It is important to realize the limitations of this approach: one cannot study

- the scalability of a local solver as there is no solver implemented. The impact of a local scheme is accounted for by the responses in spectral space.
- the effect of using an inhomogeneous Helmholtz operator. This is only possible if a local solver is implemented.

Despite these limitations, one may not underestimate the value of this method. The spectral ALADIN model offers a unique tool for such comparison studies. Working the other way around, starting from a NWP model based on a local method (e.g. ENDGame) and adding a spectral option, would be far more complex. It is, indeed, not possible to calculate in grid point space a derivative with spectral accuracy. One should need serious changes and add for example spectral transforms.
Repeating the same experiment in the global parent model IFS/ARPEGE would be complex, because then the basis functions are spherical harmonics.

5.2.2 Modifications in the ALADIN code

There is only a limited number of routines that should be changed in order to add local strategies in the ALADIN code. Table 5.2 presents an overview of the modifications needed.
It is sufficient to replace in all routines that contain spectral responses these by a more general zonal derivative (*FRESPL*), meridional derivative (*FRESPM*), and Laplacian response (*FLAPRESP*). The responses are calculated in a new routine where they are defined in function of the namelist parameters *NORDER*, which determines the order of accuracy of the FD scheme, and *LDISLOC*, which decides whether a spectral or FD scheme is used.
Appendix C shows the A-grid and Z-grid version of the routine to calculate the responses. The difference between the two approaches is limited to the definition of the Laplacian:

- **A-grid version** The A-grid version makes use of a Laplacian response that is the square of the first-order derivative response (line 124 in C.1).
- **Z-grid version** For the Z-grid version the Laplacian is calculated based on FD approximations of the second-order derivative (lines 135-165 in C.2).

For the reconstruction from D ζ to UV we use the definition of divergence and vorticity (named earlier the DZ method). The Z-grid scheme implemented here contains an asymmetry in the second-order derivative of geopotential. This could be solved by using a more complex expression for some of the first-order derivatives (see 4.3.1). However, implementing these symmetric operators in the ALADIN code turns out to be cumbersome. One needs response functions for the first-order derivative that depend both on the meridional and zonal wavenumber, as can be seen in Table 4.3. However, the current organization of the code makes that at some instants only the meridional or zonal wavenumber is known. This separation of the meridional and zonal calculations in ALADIN is a consequence of the different transforms (Legendre along meridional and Fourier along zonal direction) used in the global IFS/ARPEGE

Table 5.2: Overview of the changes implemented in the ALADIN code to study the impact of a local spatial discretization method within an otherwise unchanged ALADIN model.

routine	function	introduced changes
ald/setup/suelap.F90	set up Laplacian	repl. spectral response by *FLAPRESP*
ald/setup/suetrans.F90	set up transforms	add *YOMZFD* namelist params
arp/setup/sutrans.F90	set up transforms	add *YOMZFD* namelist params
arp/mod/yomzfd.F90	new module	containing namelist parameters *LDISLOC* and *NORDER*
tal/module/efsc_mod.F90	calc. EW deriv.	repl. spectral response by *FRESPL*
tal/module/eresp_mod.F90	new routine to calc. responses	calculation of responses *FRESPL*, *FRESPM* and *FLAPRESP*
tal/module/espnsde_mod.F90	calc. NS deriv.	repl. spectral response by *FRESPM*
tal/module/euvtvd_mod.F90	transform UV to D ζ	repl. spectral response by *FRESPM* and *FRESPL*
tal/module/ suemp_trans_preleg_mod.F90	set up distrib. envir. transforms	repl. spectral response by *FLAPRESP*
tal/module/evdtuv_mod.F90	transform D ζ to UV	repl. spectral responses by *FRESPM* and *FRESPL*

models upon which ALADIN is based. This means that implementing a symmetric scheme is technically possible but would need a reorganization of the code that is in contradiction with the modularity we strive for. We, therefore, decided to limit the test to the asymmetric Z-grid scheme and (symmetric) A-grid scheme. These options correspond to scheme 2 and scheme 4 of the previous chapter.

5.3 Finite-difference ALADIN tests in adiabatic mode

In this section we compare the results of different numerical schemes for an adiabatic ALADIN model test. The tests are undertaken with the cycle 40 release of the ALADIN code. Such a test uses the complete dynamics, but it does not include physics contributions. The model was used in a 2TL SISL configuration in hydrostatic mode with quadratic truncation in spectral space. The LAM domain consists of 97 by 97 grid points and has 46 vertical levels. The horizontal resolution is 7 km and a time step of 300 s is used.

As explained earlier, the FD methods are implemented in spectral space by modifying the responses. This means that advantages of the use of a local method, such as the use of an inhomogeneous reference state, are nonexistent in this approach. In this way, the FD runs are by definition less accurate than the spectral one. It is therefore interesting to study the differences of FD forecasts with respect to the spectral solution. The root mean square (RMS) error is used to quantify this difference. As an example, the RMS error of the 500 hPa temperature T_{500} yields:

$$RMS\left(T_{500}\right) = \sqrt{\frac{\sum_{i=1}^{N}\left[T_{500}\left(\mathbf{x}_i\right) - T_{500}^{ref}\left(\mathbf{x}_i\right)\right]^2}{N}}, \tag{5.3}$$

where the summation index i runs over all grid points and T_{500}^{ref} is the result of the current spectral ALADIN code. For this impact study only the RMS error is used, but remark that for a more thorough evaluation more scores should be computed (e.g., bias) over longer periods of time (e.g., a summer and a winter month).

We now present some findings based on Figures 5.2 and 5.3, which show the RMS error for the wind, the temperature, and the geopotential at 500 hPa (around 5.5 km height) and 925 hPa (around 800 m height), respectively. The plots present the RMS errors for second-, fourth-, sixth-, and eighth-order FD A-grid and Z-grid schemes and it also includes the results for the A-grid and Z-grid schemes combined with a spectral discretization. Confirming the earlier conclusion that a spectral spatial discretization is not influenced by the exact organization of the computations, these spectral A-grid and Z-grid experiments are indistinguishable from the spectral reference run.
Figures 5.2 and 5.3 clearly show differences between the various schemes. This proves that changing the spatial discretization does impact the results of a model run. A-grid and Z-grid schemes with the same order of accuracy result in different RMS errors. This confirms that dispersion and eigenmode properties are relevant for a real model experiment. Increasing the order of accuracy of the FD scheme yields convergence towards the spectral method.
The A-grid scheme outperforms the Z-grid one, in the sense that for the same order of accuracy the Z-grid scheme has a larger RMS error. Figures 5.2 and 5.3 show, for example, that the 4th-order A-grid approach has a smaller RMS error than the 8th-order Z-grid scheme.
We learned in the previous chapter that, despite having in theory better dispersion relations compared to the A-grid, there are two pitfalls to Z-grid schemes: asymmetry if combined with SI time discretization and distortion of the eigenvectors. The first issue

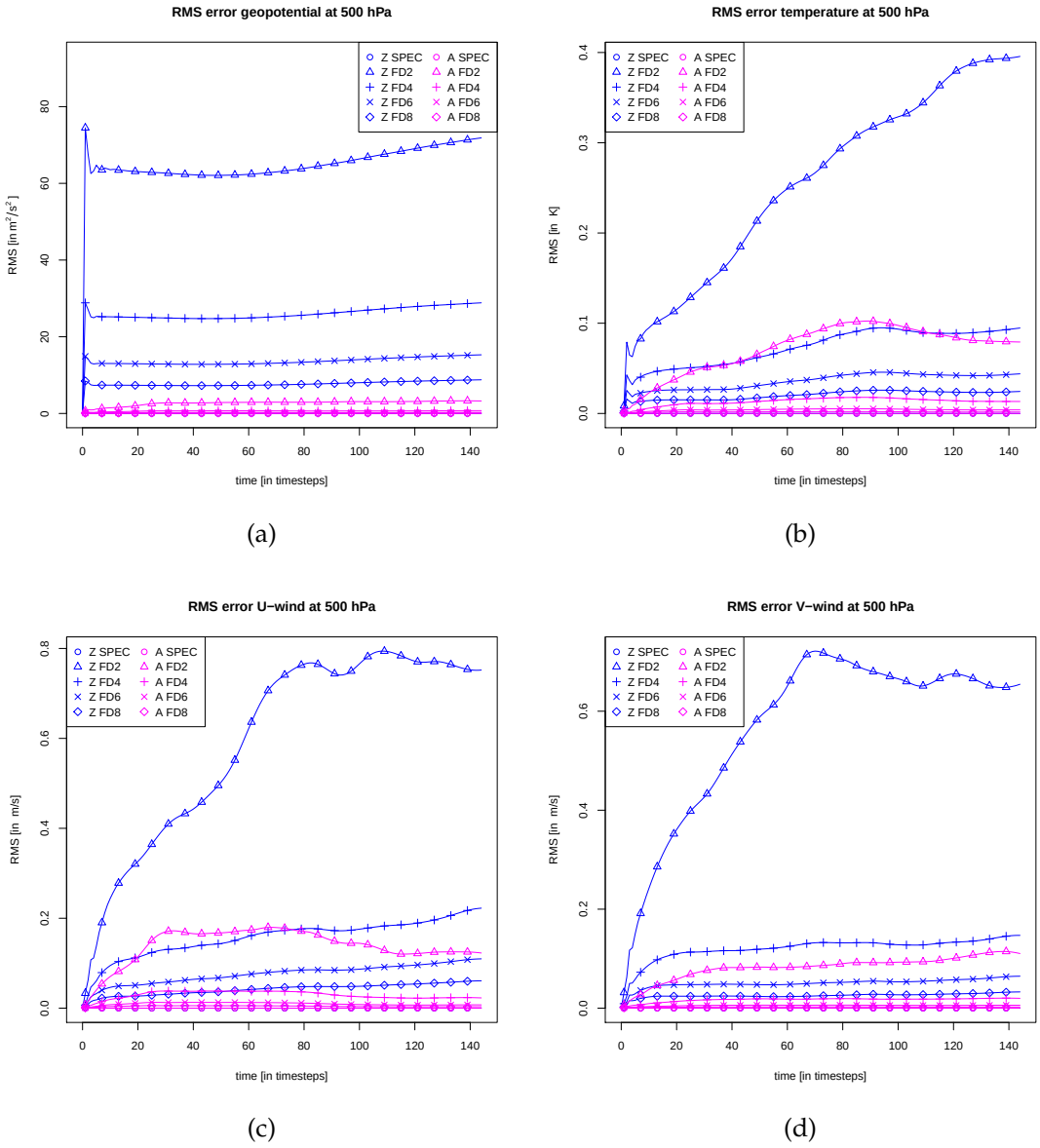
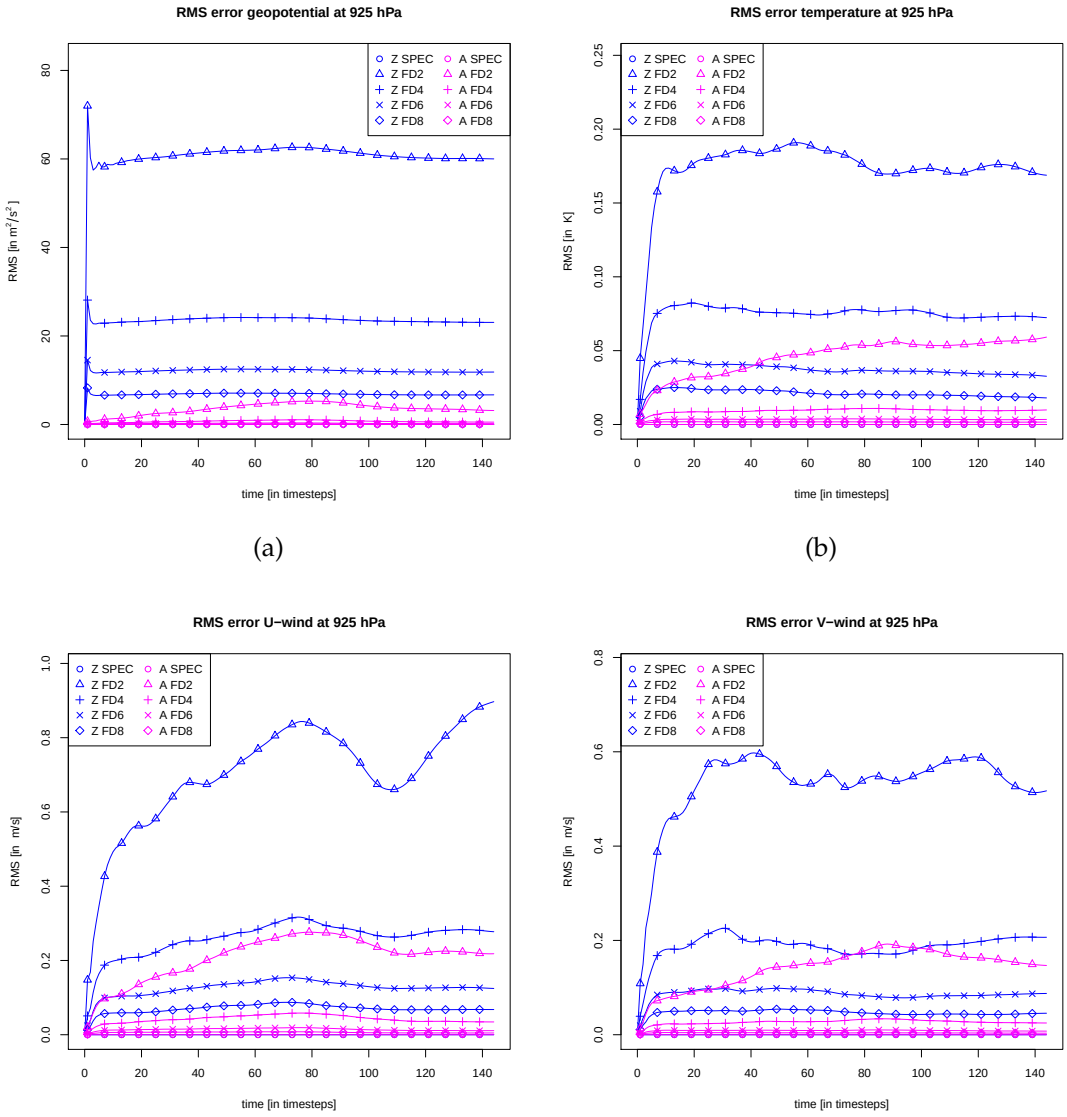


Figure 5.2: Evolution of the RMS error during the first 12 hours of an adiabatic ALADIN forecast: geopotential (a), temperature (b), zonal wind (c) and meridional wind (d) at 500 hPa. Magenta and blue represent the runs with the asymmetric SISL Z-grid and the A-grid method, respectively. The symbols define the order of accuracy.



(a) (b)

(c) (d)

Figure 5.3: Idem as Figure 5.2 but at 925 hPa.

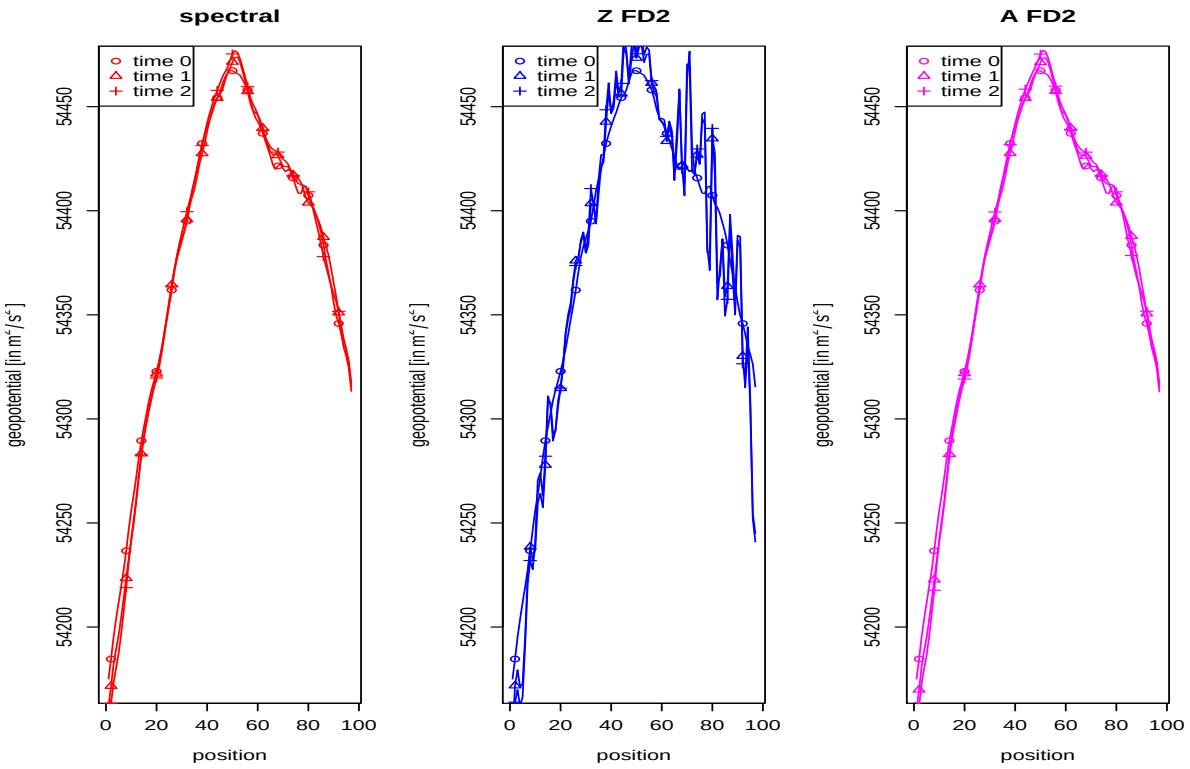
could be solved by making the scheme symmetric at the cost of some extra complexity, the second is intrinsically linked to the Z-grid scheme. Determining which of these two forms the main contribution to the RMS error of the asymmetric Z-grid scheme that was used in this study could be done by including the symmetric Z-grid scheme. But, as explained earlier, this falls out of the scope of this work. However, a careful look at the fingerprint of the RMS evolution reveals the main contribution.

The RMS error of Z-grid schemes shows a peak at the start of the experiment, especially for geopotential. Moreover, the maximal RMS error for the geopotential field is found after the first time step. This is clearly a fingerprint of the misprojection of the initial state onto the different modi, as explained and illustrated in chapter 4. For the SWE tests this misrepresentation only spoiled some fields. For example, the DZ-based scheme gave poor results for the wind field in the geostrophic adjustment test while there was no problem concerning the geopotential (Figure 4.10). In a complete model there is a strong coupling between the different fields causing a spill over of the short scale noise to other fields, causing a sharp increase of the RMS error for multiple fields.

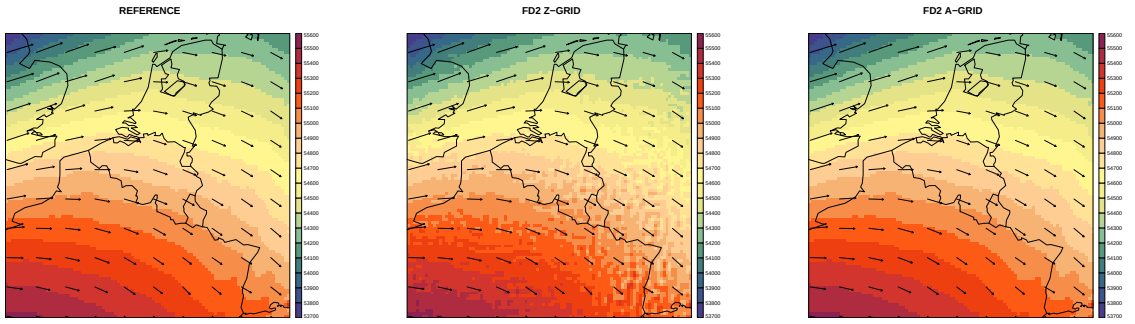
More evidence for the eigenvector attribution is found by studying the geopotential field during the first time steps. Figure 5.4 (a) compares 1D crossections of the 500 hPa geopotential during the first time steps for the second-order FD A-grid, Z-grid, and spectral reference run. For the A-grid and reference run one merely notices an evolution during the first time steps, whereas for the Z-grid method short scale noise pops up in the geopotential field after the first time step. This agrees with the analysis presented in section 4.4. Figures 5.4 (b), (c), and (d) show the 2D geopotential field at 500 hPa after the first time step. The noise of the Z-grid scheme is again clearly visible. It turns out that this noise remains present in the Z-grid scheme during the complete run, as illustrated by Figure 5.5. As the distorted eigenvector decomposition turns out to be at the origin of the poor Z-grid results, a symmetric Z-grid scheme, which was not included in our tests, will not behave better.

It would be interesting to repeat the previous tests with an initialization procedure, for example digital filter initialization (DFI) [62]. The DFI procedure removes large amplitude IGWs from the initial state and may therefore remove or weaken the noise that pops up with the Z-grid method.

We conclude that the known dispersion problems of the FD A-grid approach are controlled in this ALADIN test, probably due to the truncation, SL advection, and other diffusive processes, whereas the FD Z-grid eigenvector problem does suffer from the eigenvector decomposition problem.

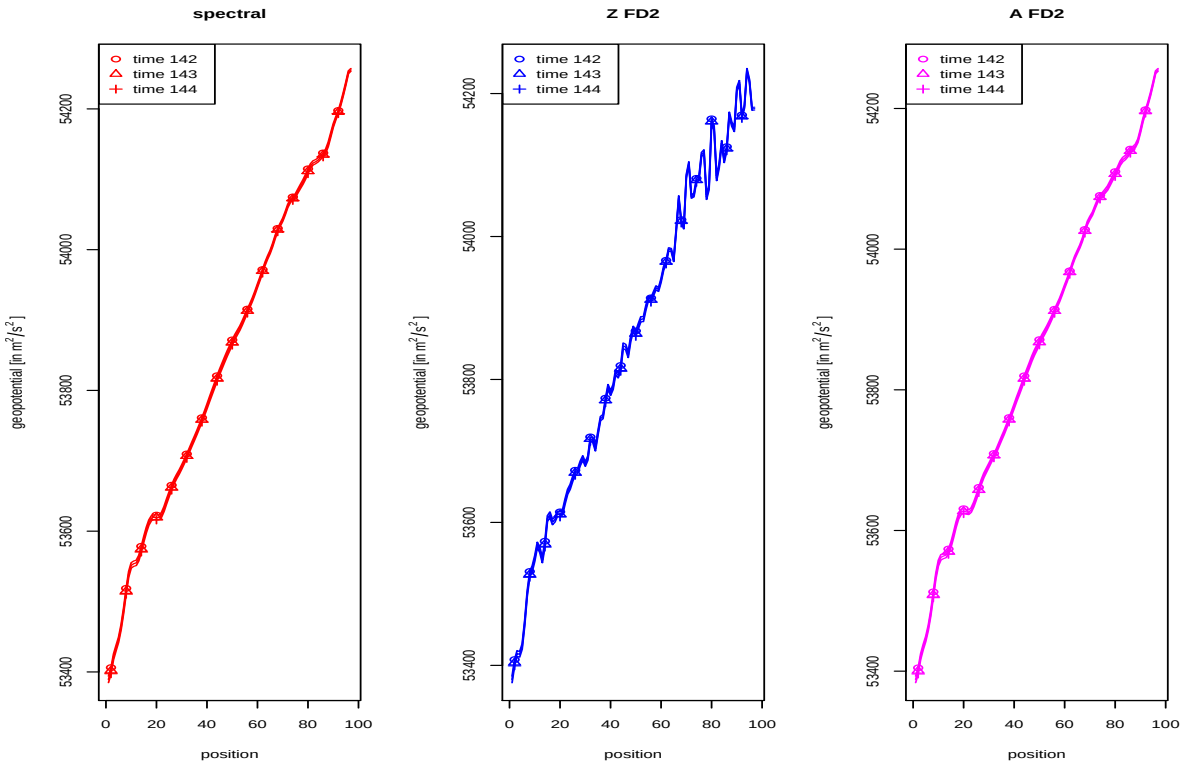


(a)

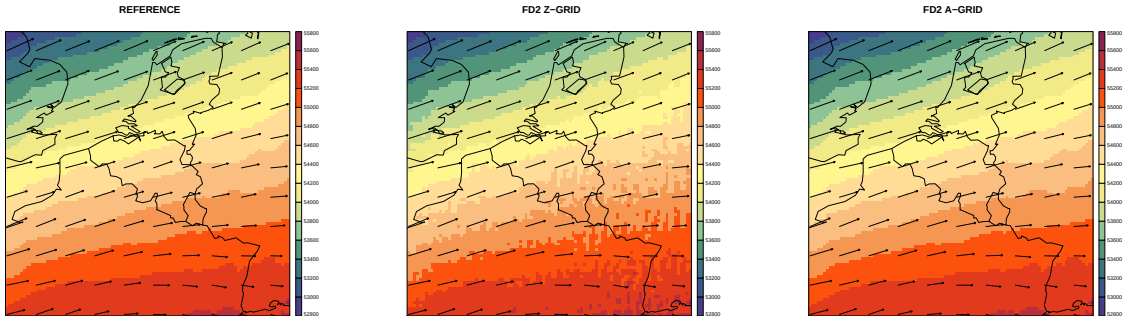


(b) (c) (d)

Figure 5.4: (a) Evolution of the geopotential at 500 hPa along a crossection of the domain during the first time steps for an adiabatic ALADIN run. From left to right: spectral reference, second-order FD Z-grid, and second-order FD A-grid. Geopotential field after the first time step for the reference run (b), the second-order FD Z-grid (c), and second-order FD A-grid (d).



(a)



(b) (c) (d)

Figure 5.5: Idem as Figure 5.4, but for the final time steps.

5.4 Finite-difference ALADIN tests with ALARO physics

One should include physics contributions to make the model tests even more realistic. In this section we will repeat the previous test but include the ALARO physics package. In this way, the experiment closely resembles an operational weather forecast. Detailed information about the convection, turbulence, and other parameterizations used in ALARO can be found in [32] and [46].

The physics influences the shortest scales of the wave spectrum. Some parameterizations excite short scale waves (e.g., deep convection) while others have a damping effect (e.g., turbulence). What may be the impact of including ALARO physics on the previous, adiabatic conclusions? On the one hand, these parameterizations add extra nonlinearity to the model. Some parameterizations are even based on stepwise functions, for instance, precipitation is formed if the relative humidity exceeds a certain threshold. This may result in an increased spread between the different experiments. On the other hand, we know that for the experiments in the previous section the differences were originating from the shortest scale part of the spectrum. In this ALARO experiment we add an extra piece of code, identical for all experiments, mainly impacting these short scales. One may reason that the shortest scales are no longer monopolized by the diverging dynamics and that adding physics is a way to control inappropriate wave behavior from the dynamics.

Figures 5.6 and 5.7 reproduce the Figures 5.2 and 5.3, but this time for an ALADIN run with ALARO physics. All settings of the dynamics were identical to the adiabatic experiment of section 5.3. In general, the conclusion of the adiabatic runs still hold, the eigenvector problem is for example still present in the Z-grid approach. However, a careful comparison of the adiabatic and ALARO RMS errors suggests that in general the deviation with respect to the reference run is slightly reduced. This may suggest that the physics are able to control partly the distortions introduced by the dynamics at the shortest scales.

Remark that the spectral A-grid and Z-grid runs, which are in theory identical to the reference run, do no longer give identical results. It is known that due to reprogramming some routines roundoff errors are introduced. In contrast to the adiabatic runs, the nonlinearity of the physics causes these tiny differences to increase to significant RMS errors. This is illustrated in Figure 5.8, which shows for the A-grid the RMS errors of some 500 hPa fields. The 4th-, 6th-, and 8th-order FD A-grid schemes result in RMS errors comparable to the RMS error of the spectral A-grid scheme, which differs

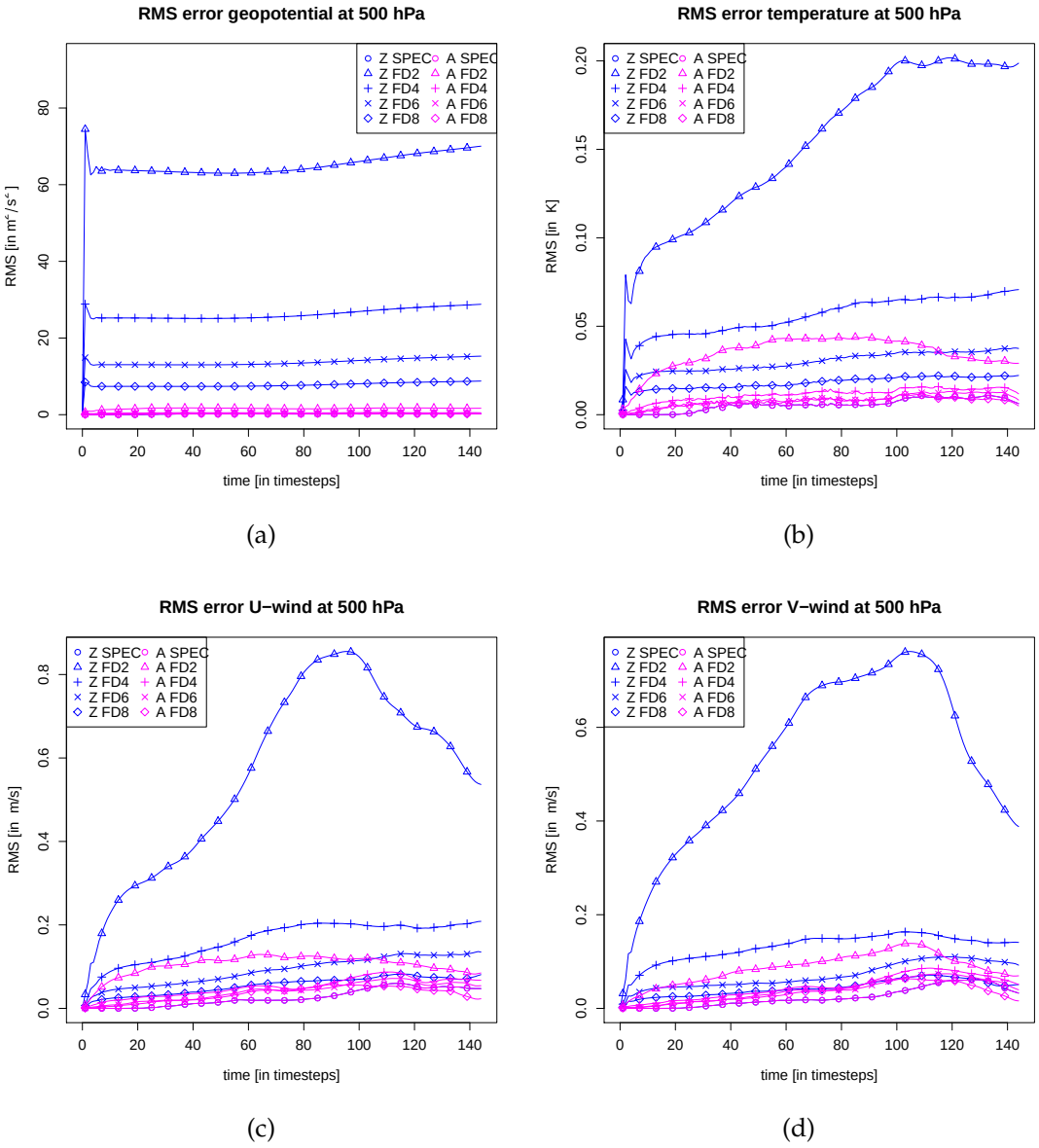
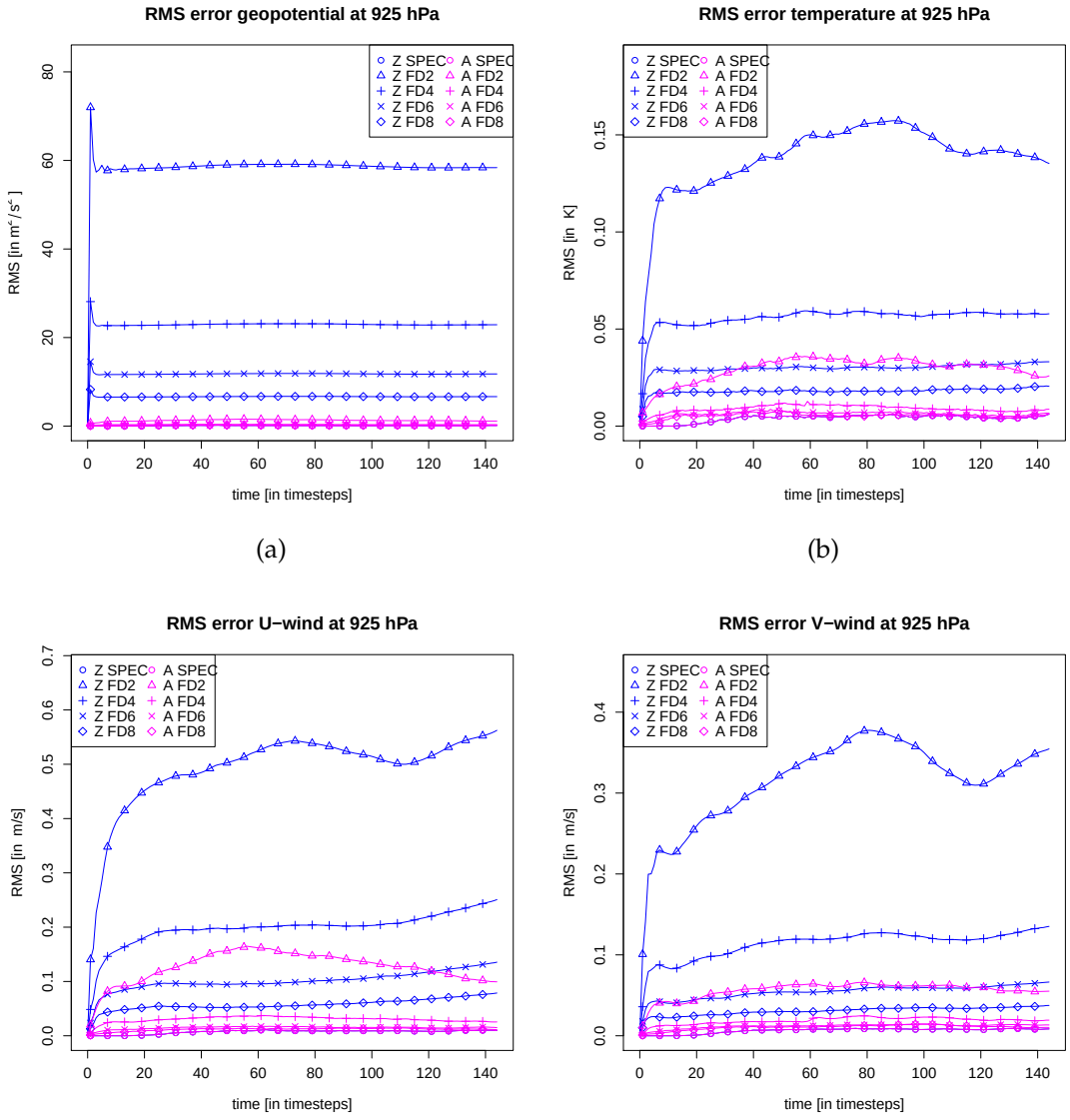


Figure 5.6: Evolution of the RMS error during the first 12 hours of a ALADIN forecast with ALARO physics: 500 hPa geopotential (a), 500 hPa temperature (b), 500 hPa zonal wind (c), and 500 hPa meridional wind (d). Magenta and blue represent the runs with the asymmetric SISL Z-grid, respectively the A-grid method. The symbols define the order of accuracy.



(a) (b)

(c) (d)

Figure 5.7: Idem as Figure 5.6 but at 925 hPa.

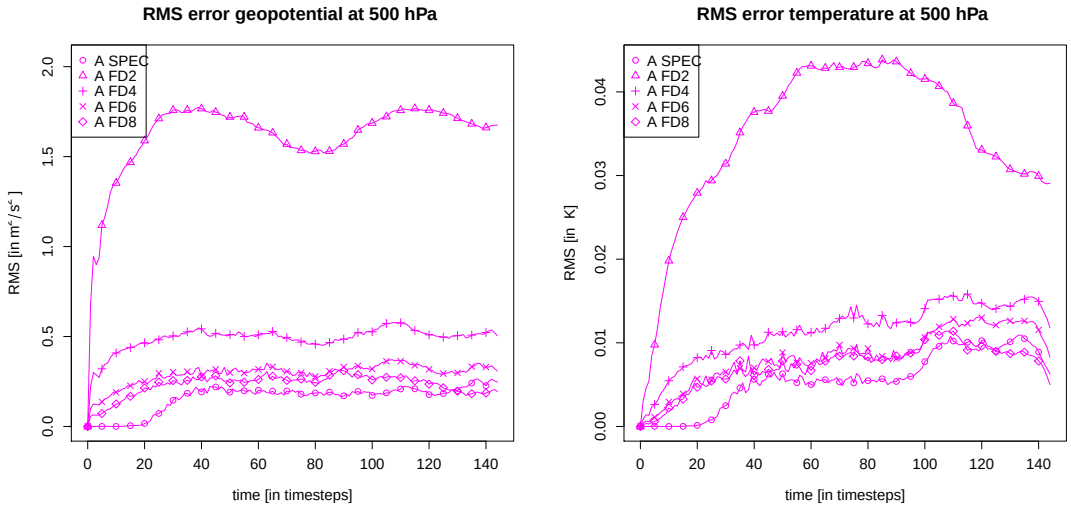
from the current spectral scheme only by some roundoff errors. This suggests that the use of high-order FD A-grid schemes results in forecasts close to the current spectral ALADIN model.

5.5 Conclusion

This chapter opens with a discussion on the implementation of local discretization approaches within the ALADIN model. It is shown that the modular approach guarantees that the majority of the current set-up can be reused. The main challenge consists of developing an efficient, scalable sparse matrix solver to solve local Helmholtz problems. This solver should invert very large, asymmetric matrices in order to have the possibility to include inhomogeneous terms in the Helmholtz problem. Both in- and outside of the NWP community, there is a lot of knowledge about iterative solvers for this kind of sparse matrix problems. Implementing a local solver therefore seems viable, but falls out of the scope of this thesis.
Such a solver is not needed to study the scientific impact of a change in spatial discretization. The ALADIN framework can be used to test the impact of different horizontal spatial discretization strategies on a full NWP model with all its complexities. FD methods are emulated by replacing the spectral responses by the corresponding FD responses. With an absolute minimum of modifications we are able to mimic FD schemes and study the differences with the current spectral method. This makes the ALADIN framework a unique and interesting testbed to study the scientific impact of spatial discretization methods.

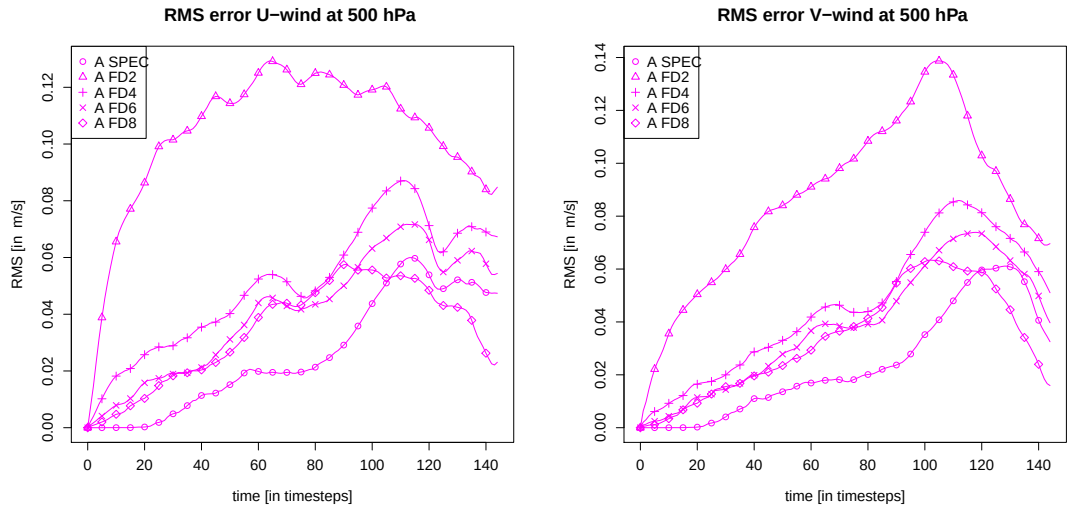
The experiments confirm that the choice of the spatial discretization impacts the results of a model run. Both in tests with and without ALARO physics, the Z-grid scheme suffers from substantial RMS errors and this from the first time step on. This agrees well with the fingerprint of the eigenvector problem that was diagnosed in chapter 4. One should note that the test was based on the asymmetric Z-grid scheme. Switching to a more complex symmetric Z-grid scheme will not solve the problem because the symmetric scheme still suffers from the eigenvector distortion. However, it would be interesting to test whether a DFI procedure could reduce this noise. The A-grid methods deviate only little from the spectral reference especially if one uses higher-order FD. The well-known negative group velocities of A-grid methods do not seem to pose a problem within a complete NWP model, at least not for the time step and resolution of this test.

At this point, it is interesting to look back to chapter 3 where it was concluded that



(a)

(b)



(c)

(d)

Figure 5.8: Same as Figure 5.6, but now only the A-grid schemes are plotted.

both A-grid and Z-grid were potential local spatial discretization approaches. The Z-grid displayed excellent IGW dispersion properties and it was therefore decided to spend an extra study on SISL Z-grid schemes. Apart from the danger of spoiling the IGW dispersion by introducing asymmetries in the implicit discretization, this study revealed that the local Z-grid schemes have aberrant eigenmodes at the short scale end of the spectrum. It seems that leaving the spectral has anyway a negative impact on the quality of the waves permitted by the discretized algorithm, independent of the exact method used. Moreover, ALADIN tests with and without ALARO physics suggest that the decomposition problem of the Z-grid methodology has a larger impact on the runs than the A-grid dispersion issues. It, therefore, seems reasonable to conclude that a higher-order A-grid method may be the most promising local spatial discretization option.

Summary and outlook

6

In this final chapter a summary, containing the most important findings of this thesis, is presented and an outlook for future research is given.

6.1 Summary

Modeling the evolution of our atmosphere accurately is beneficial for society. The groundbreaking work of Vilhelm Bjerkness [10] and Lewis Fry Richardson [85] laid the foundations of numerical weather forecasting. Last decades, the accuracy of the forecasts has been improved strongly, mainly by the increase in available computing power and high-quality observations, the improvements in data assimilation, and a better understanding of the atmospheric processes and numerical methods. The ambition for the modeling consortia consists of further improving the quality of the forecasts by using the resources optimally. With this study we hope to add a small contribution to the achievement of this huge challenge.

Chapter 1 presented a historical introduction on numerical weather prediction (NWP) and a non-exhaustive overview of the different continuous equations systems and discretization schemes that are used in today's NWP models. It was explained that choices made in a NWP model should be seen within the context where the model is developed in; this framework imposes constraints that are beyond the control of the model developer.
This study is situated within the context of the ALADIN (Aire Limiteé Adaptation dynamique Développement International) model, which is used operationally in 16 countries. The ALADIN numerics combine the highly accurate spectral horizontal spatial discretization with a semi-implicit semi-Lagrangian (SISL) time discretization, which permits the use of long timesteps while staying stable. Due to the diagonal character of derivative operations in spectral space, the Helmholtz problem that results from the SI discretization is solved trivially in spectral space. Because some of the computations, for example the subgrid physics parameterizations, are evaluated in grid points, transforms between spectral and grid point space are needed every timestep. The global

character of the spectral basis functions makes that these transforms need global communication. Combined with the trend towards massively parallel supercomputing, where communication seems to become a bottleneck, this could undermine the scalability of the spectral approach. Apart from its potential scalability problem, there is a second disadvantage: a spectral method can not include horizontally inhomogeneous terms in the Helmholtz equation. Therefore, the nonlinear residual terms, which are treated with an explicit timestepping scheme, are substantial and this may pose problems for the stability of the scheme.
As local spatial discretization methods fit within the two previous limitations, this thesis investigated whether and how one could implement local methods, for example a finite difference (FD) scheme, within the current ALADIN numerics.

Model developments need to obey external constraints that are imposed on the NWP model. Therefore, chapter 2 gave an overview of three constraints that are relevant for the ALADIN numerics:

1. The goal of the ALADIN model is to deliver timely high-resolution mesoscale forecasts on limited area domains.
2. Code is developed in a modular way such that the impact on the engineering applications and products of the model, on other parts of the model, and on the users is minimal. For this thesis the modularity constraint is translated into two recommendations: maintain the timestep organization and stay on a collocation or unstaggered grid, where all variables are defined in the same grid points.
3. The model must exploit the increased computing power.

Atmospheric motion can be decomposed into different sorts of waves. Some, for instance Rossby waves, have large meteorological relevance, whereas others, such as acoustic waves, are not of any meteorological interest. Inertia-gravity waves (IGWs) have a position in between. They are of direct importance for some mesoscale meteorological phenomena. However, at synoptic scales they only have an indirect relevance. At these scales the atmosphere is always striving towards a geostrophic balance between the geopotential and velocity fields. Imbalances to this geostrophic adjustment are restored by radiating away IGWs.

As illustrated in chapter 3, discretization in space and time causes the numerical atmospheric waves to be different from their analytical counterparts. By its high order of accuracy, the spectral discretization method does represent all waves accurately. However, a local discretization method on an unstaggered grid (A-grid) results in inappropriate geostrophic adjustment for the shortest scale IGWs. Therefore, many models

make use of a staggered grid where different variables are defined in different grid points. Another solution proposed in literature, called the Z-grid method, consists of reformulating the equations from wind components to divergence and vorticity. In this way, one can recover appropriate IGW propagation while staying on an unstaggered grid, which is one of the constraints for the model. Therefore, chapter 4 presented a more detailed study of the Z-grid method within the SISL ALADIN framework.

The detailed SISL Z-grid study in chapter 4 revealed two important points of attention:

- Combining the Z-grid approach with a SI time discretization results in a scheme containing asymmetries between spatial operators in the explicit and implicit part of the computations. Analysis and toy model tests based on the linear SWE show that these asymmetries result in unphysical IGW dispersion that annihilates the appropriate Z-grid properties. Local SISL Z-grid schemes that were proposed in literature suffer from this problem.
- The eigenvector decomposition of a Z-grid scheme is incorrect for the shortest scale part of the spectrum, independent of the time discretization. Short scale noise in the wind fields for Z-grid geostrophic adjustment tests confirm this analysis.

By modifying some of the operators in the explicit part of the SISL computations, symmetry can be restored. However, the eigenvector decomposition is inherent to the Z-grid method and can not be solved easily by tweaking the scheme or its operators. Adding non-linear terms, such as advection, or truncation to the SWE toy model makes that the inappropriate wave behavior becomes less obvious. Therefore, one could wonder how relevant dispersion properties are within the context of a complete NWP model.
Finally, chapter 4 illustrated the possibility of local schemes to include inhomogeneous terms in the Helmholtz solver. By doing so, the non-linear residuals, which are a source of potential instabilities and are treated in an explicit way, are reduced.

Chapter 5 explained that introducing a SISL scheme based on a local spatial discretization method fits well within the current ALADIN timestep organization. The main change would be the introduction of a sparse matrix solver for the Helmholtz problem to replace the current diagonal problem in spectral space. For the development of such a solver one could make use of the extensive experience in different branches of science about iterative solvers for large, sparse matrix systems on massively parallel machines.

Despite the fact that the implementation of such a solver falls out of the scope of this

thesis, chapter 5 presented results of FD runs undertaken with the ALADIN model. Adding the FD wave responses to the current spectral code is sufficient to mimic the use of FD schemes in ALADIN. This methodology provides a powerful testbed to investigate the scientific impact of different spatial discretization methods.

One high-resolution LAM forecast with quadratic truncation was repeated for different spatial discretization schemes. The standard spectral run was considered as the reference. Some conclusions:

- The FD schemes based on the asymmetric Z-grid approach had larger differences with the reference run than the A-grid based schemes. A detailed look clearly showed the fingerprint of the distorted projection of the eigenvectors by the Z-grid scheme. As the decomposition is crucial during the first timesteps, a proper initialization (e.g., DFI) may weaken or remove this problem.
- There was no fingerprint found of the negative group velocity of the A-grid schemes.
- By increasing the order of accuracy of the FD scheme, the local schemes converged toward the reference run.

These conclusions were valid for ALADIN tests both in adiabatic and ALARO mode.

Let us come back to the question that formed the basis of this thesis: *Can we use within the current spectral SISL ALADIN model a local horizontal spatial discretization scheme and how to do this?*.
This thesis provides arguments that a local solver can be added to the ALADIN framework while retaining most of the current code organization. FD spatial discretization methods based on the Z-grid approach suffer from an eigenmode decomposition problem, which mainly manifests itself during the first timesteps. Similar FD tests were undertaken within an A-grid approach and no fingerprint of the spurious waves that are diagnosed in analytical A-grid tests was found. The A-grid approach combined with fourth- or higher-order FD spatial discretization yields results close to the spectral experiments for ALARO tests. Therefore, higher-order A-grid methods are a promising candidate for a modular implementation of local schemes within ALADIN.

6.2 Outlook

This thesis presented a general study about the impact of local spatial discretization choices for the currently spectral ALADIN model. The implementation of a local ALADIN solver was out of the scope of this work. However, this work might bring such

an implementation one step closer. In what follows, two interesting topics for future research are put forward.

Primarily, it would be interesting to undertake more studies within the ALADIN testbed to further investigate the impact of the responses of local methods. The relevance of this research is far broader than the ALADIN community as many modeling groups currently use local spatial discretization methods. Often they implement approaches more expensive than the straightforward A-grid method (e.g., grid staggering, mixed FE,...) to avoid A-grid dispersion problems. Studying further the relevance of dispersion issues in a full model context would therefore be valuable. Some potential ideas:

- currently the tests of Chapter 5 are done for one forecast. Repeating this test for different cases on different domains with different resolutions would make the conclusions more robust.
- study the influence of truncation. The presented ALADIN tests were done with quadratic truncation but what would happen without truncation or with cubic truncation?
- do some tests with a digital filter initialization, known as DFI [62], to verify whether this could reduce the eigenvector decomposition problem of the Z-grid scheme.
- study a scheme where part of the computations are done in a spectral way (e.g., the Helmholtz equation) and part in a local way (e.g., the derivatives in the explicit right hand sides)? Such a scheme is sometimes proposed as a compromise between spectral and local approaches but according to our analysis it results in an asymmetric approach...

The implementation of a local solver is another research priority. As a first step, an existing iterative solver (Krylov or multigrid) should be applied on the problem as it is currently defined in the hydrostatic version of ALADIN. If one keeps everything identical (same LBC approach, biperiodize the fields, diffusion in spectral space,...) and solves the Helmholtz problem with a local solver, one should be able to reproduce the FD A-grid and Z-grid results that were found in this thesis by modifying the responses. Once these results are reproduced by a local solver one could start to

1. move step by step to a more realistic local A-grid SISL scheme by investigating a local way to implement diffusion, reformulating the LBCs, adding inhomogeneous terms to the Helmholtz solver,...
2. test the scalability of the local solver on different architectures.

Samenvatting

7

Het nauwkeurig modelleren van onze atmosfeer is van groot belang voor de maatschappij. De vergelijkingen die de evolutie van de atmosfeer beschrijven hebben geen analytische oplossingen en dienen daarom numeriek opgelost te worden. De publicaties van Vilhelm Bjerkness [10] en Lewis Fry Richardson [85] legden de basis voor zulke numerieke weersvoorspellingen. Door de toename aan beschikbare rekenkracht en hoogwaardige observaties en een beter begrip van meteorologische processen en numerieke methodes zijn de voorspellingen van weermodellen de afgelopen decennia veel nauwkeuriger geworden. Onderzoekers betrokken bij weermodellering hebben de ambitie om de verder toenemende rekenkracht optimaal aan te wenden om de voorspellingen te verbeteren. Deze thesis hoopt hiertoe een kleine bijdrage te kunnen leveren.

Hoofdstuk 1 schetste de historische ontwikkelingen van numerieke weermodellering. Daarnaast werd een overzicht gegeven van vergelijkingen en numerieke methodes die vaak gebruikt worden in hedendaagse weermodellen. Er werd uitgelegd dat de ontwikkeling van een weermodel niet los gezien kan worden van het kader waarin het ontwikkeld wordt. Dit kader legt een aantal externe voorwaarden op waaraan het weermodel dient te voldoen.
Deze thesis kadert binnen de context van het ALADIN model dat momenteel operationeel gebruikt wordt in 16 landen, waaronder België. De numerieke oplossingsmethode van het ALADIN model combineert de spectrale methode voor de ruimtelijke discretisatie met een semi-impliciete en semi-Lagrangiaanse (SISL) tijdsdiscretisatie. De keuze voor een SISL tijdsdiscretisatie laat grote tijdsstappen toe zonder dat het model instabiel wordt. De spectrale methode benadert alle meteorologische variabelen (bv. wind, temperatuur,...) als een som van functies die over het volledige domein gedefinieerd zijn. Het globale karakter van deze basisfuncties maakt dat de spectrale methode zeer nauwkeurige berekeningen mogelijk maakt in spectrale ruimte.
Sommige berekeningen dienen echter plaats te vinden in de roosterpunten. Het gaat hierbij bijvoorbeeld om het in rekening brengen van processen die te kleinschalig zijn om expliciet op het rooster voorgesteld te kunnen worden. Hierdoor dient er elke

tijdsstap een transformatie uitgevoerd te worden van de variabelen gedefinieerd in de roosterpunten naar de spectrale ruimte, en terug. Door het globale karakter van de spectrale basisfuncties, zijn er voor deze transformatie functiewaarden uit het volledige domein nodig. De nood aan globale communicatie maakt de schaalbaarheid van spectrale methodes op de toekomstige supercomputers onzeker. In de toekomst zullen deze immers uit miljoenen processoren bestaan die allen parallel werken en waar communicatie tussen de processoren relatief traag verloopt. Daarnaast laat een spectrale methode niet toe om horizontaal inhomogene termen op te nemen in de SI tijdsdiscretisatie en dus Helmholtzvergelijking. Dit betekent dat zulke termen op een expliciete manier behandeld moeten worden wat de stabiliteit van het model in het gedrang kan brengen.
Lokale discretisatiemethodes, zoals eindige verschillen (EV), kunnen soelaas bieden voor deze twee beperkingen van de spectrale methode. Daarom onderzoekt deze thesis of en hoe het mogelijk is om een lokale methode voor de horizontale ruimtelijke discretisatie in te voeren binnen de ALADIN context.

Zoals eerder gezegd, dienen modellen een aantal extern opgelegde voorwaarden in acht te nemen. In hoofdstuk 2 werden er drie besproken die relevant zijn voor deze studie:

1. Het ALADIN model heeft als doel om binnen de voorziene tijd voor een beperkt gebied hoge resolutie mesoschaal weersvoorspellingen te leveren.
2. De code wordt ontwikkeld op een modulaire manier zodat de gevolgen van veranderingen voor de andere delen van het model, de toepassingen en de gebruikers minimaal zijn.
3. Het model dient optimaal gebruik te maken van de rekenkracht die het ter beschikking heeft.

De toestand van de atmosfeer kan ontbonden worden in verschillende soorten golven. Sommigen, zoals bijvoorbeeld Rossby golven, zijn van zeer groot belang voor weersvoorspellingen. Maar anderen, zoals geluidsgolven, hebben geen enkele relevantie voor weermodellering. Inertie-gravitatie golven (IGG) hebben een positie tussenin. Zij zijn van rechtstreeks belang voor mesoschaal modellering maar op grotere schaal zijn ze slechts indirect relevant. Op de synoptische schaal streeft de atmosfeer naar een evenwicht tussen druk en wind, het zogenaamde geostrofische evenwicht. Verstoringen van dit evenwicht worden hersteld door excitatie en propagatie van IGG.

Zoals uitvoerig besproken in hoofdstuk 3 zal ruimtelijke en tijdsdiscretisatie de eigenschappen van de numerieke golven in een weersvoorspelling beïnvloeden. Door zijn

hoge nauwkeurigheid is de spectrale ruimtelijke discretisatie in staat om de analytische golfoplossingen te reproduceren. Maar dit is niet langer het geval als lokale methodes, zoals EV ruimtelijke discretisatie, gebruikt worden. In de literatuur wordt uitvoerig beschreven hoe lokale ruimtelijke discretisatiemethodes tot foutieve propagatie van de kortste IGG leidt. Dit gebeurt als de vergelijkingen geformuleerd zijn in functie van wind en opgelost worden op een collocatierooster, waar alle functies in dezelfde roosterpunten gedefinieerd zijn. Deze methode wordt de A-rooster methode genoemd. Daarom maken heel wat weermodellen gebruik van een rooster waar verschillende variabelen in verschillende roosterpunten gedefinieërd zijn, bijvoorbeeld de C-rooster methode. Een andere oplossing om gepaste IGG propagatie te bekomen, bestaat erin om op een collocatierooster te blijven, wat de modulariteit ten goede komt daar het ALADIN model momenteel zulk rooster gebruikt, maar de vergelijkingen te herschrijven in functie van vorticiteit en divergentie. Het toepassen van deze zogeheten Z-grid methode binnen de SISL ALADIN context werd in meer detail besproken in hoofdstuk 4.

De gedetailleerde Z-rooster studie van hoofdstuk 4 leidde tot twee belangrijke aandachtspunten:

- De combinatie van een Z-rooster methode met een SI tijdsdiscretisatie kan aanleiding geven tot een asymmetrie tussen de ruimtelijke operatoren in het expliciete en impliciete gedeelte van het schema. Analyse en academische modeltests tonen aan dat zulke asymmetrieën leiden tot een niet-fysische IGG dispersierelatie en zo de voordelen van de Z-rooster methode teniet doen. Lokale SISL Z-rooster schema's die in het verleden voorgesteld werden, vertoonden dergelijke asymmetrieën.
- De eigenvectoren van een Z-rooster schema blijken afwijkingen te vertonen voor de kortste golven en dit onafhankelijk van de tijdsdiscretisatie. In numerieke academische tests werd inderdaad kortgolvige ruis teruggevonden bij Z-rooster schema's.

Door sommige operatoren in het expliciete gedeelte van de berekeningen aan te passen, is het mogelijk om de symmetrie te herstellen. Maar het probleem van de eigenvectoren is inherent aan de Z-rooster methode en kan niet verholpen worden door operatoren te veranderen. Als de numerieke tests ook niet-lineaire termen bevaten, zoals bijvoorbeeld advectie, of als het spectrum getrunceerd wordt, dan komen de IGG problemen minder duidelijk naar voren. Daarom stelt zich de vraag of zulke misrepresentaties eigenlijk wel relevant zijn voor een operationeel weermodel.
Tenslotte werd in hoofdstuk 4 aangetoond hoe een lokale ruimtelijke discretisatiemethode toelaat om inhomogene termen op te nemen in de Helmholtzvergelijking.

Hoofdstuk 5 legde uit dat het inpassen van een lokale Z-rooster methode binnen de SISL ALADIN tijdsstaporganisatie mogelijk is. De grootste verandering die hiervoor nodig is, bestaat uit de implementatie van een code die op een efficiente manier ijle matrices, die slechts een beperkt aantal elementen verschillend van 0 bevatten, kan inverteren. Deze zal de huidige Helmholtz oplossingsmethode in de spectrale ruimte vervangen. Voor deze ontwikkeling kan gebruik gemaakt worden van de ervaring die in veel onderzoeksdomeinen opgebouwd is betreffende het oplossen van zulke matrix-problemen op massief parallelle supercomputers.

Ondanks het feit dat dergelijke implementatie buiten het doel van deze thesis valt, stelden we in hoofdstuk 5 toch voorspellingen voor met het ALADIN model gebaseerd op een EV discretisatie. Dit was mogelijk omdat men in spectrale ruimte de responsfuncties van de ruimtelijke afgeleiden zo kan aanpassen dat men resultaten bekomt identiek als zou een EV methode gebruikt zijn. Deze methodologie biedt een unieke testomgeving om de wetenschappelijke impact van verschillende ruimtelijke discretisatiemethodes op een eenvoudige manier te analyseren.

Een weersvoorspelling met het ALADIN model werd verschillende malen herhaald voor verschillende keuzes van de ruimtelijke discretisatie. Het resultaat van de huidige spectrale methode werd als referentie beschouwd. Enkele vaststellingen:

- De EV schema's die gebaseerd waren op asymmetrische Z-rooster methodes vertoonden duidelijk grotere afwijkingen ten opzichte van de referentie dan die met de A-rooster methode. Bij nader inzien bleek de verkeerde voorstelling van de eigenvectoren bij een Z-rooster methode aan de basis te liggen van deze afwijking. Omdat dit vooral opspeelt tijdens de eerste tijdsstappen van de voorspelling, zou het kunnen dat een gepaste initialisatie dit probleem kan verhelpen of inperken.
- Er werd geen enkele aanwijzing gevonden van de IGG problemen die een A-rooster methode kenmerken.
- Bij het verhogen van de orde van nauwkeurigheid van de EV methode, bleken deze voorspellingen te convergeren naar de spectrale referentievoorspelling.

Deze vaststellingen waren geldig zowel voor adiabatische ALADIN tests als voor voorspellingen waarbij de ALARO-parameterisaties gebruikt werden.

Tot slot komen we nog eens terug op de onderzoeksvraag waarrond deze thesis opgebouwd is: *Kunnen we binnen de huidige spectrale SISL ALADIN context een lokale ruimtelijke*

discretisatiemethode introduceren en hoe zou dit dan in zijn werk moeten gaan?
Deze thesis geeft aan dat zulke lokale methodes kunnen gebruikt worden binnen ALADIN met behoud van het grootste gedeelte van het huidige model. Een EV ruimtelijke discretisatie gebaseerd op een Z-rooster methode heeft een eigenvector probleem dat vooral naar voren komt bij de aanvang van een voorspelling. Gelijkaardige EV tests maar met een A-rooster tonen dat de analytisch verwachtte IGG dispersieproblemen niet teruggevonden worden in een weersvoorspelling. De A-rooster methode in combinatie met vierde of hogere-orde EV afgeleides leidt tot resultaten zeer dicht bij de huidige spectrale resultaten. Dit blijkt dan ook een veelbelovende kandidaat te zijn voor modulaire implementaties van lokale schema's binnen ALADIN.

Linear finite-element discretization

A

In this appendix, the linear finite-element (FE) approach is used to illustrate the sparse character of Galerkin methods that are based on local basis functions. Some of the 1D linear FE basis functions $\phi_i(x)$ are shown in Figure A.1. Remark that the basis functions overlap only with their closest neighbors.
Let us retake the matrix formulation of a constant advection problem if a Galerkin method is used:

$$\mathbf{M} \times \frac{d\mathbf{U}}{dt} = \mathbf{D} \times \mathbf{U}, \tag{A.1}$$

with

$$M_{ji} = \int_\Omega \phi_i(x)\,\phi_j(x)\,dx \text{ and} \tag{A.2}$$

$$D_{ji} = \int_\Omega \frac{d\phi_i}{dx}(x)\,\phi_j(x)\,. \tag{A.3}$$

In the previous, $\mathbf{M}$ is the $N \times N$ mass matrix, $\mathbf{D}$ is the $N \times N$ differentiation matrix, and $\mathbf{U}$ is the $N \times 1$ column vector containing the coefficients of the basis functions. Unlike the spectral method, the linear FE method does not consist of orthogonal basis functions. However, the overlap of the basis function ϕ_i and its derivative $\frac{\partial \phi_i}{\partial x}$ is limited to the neighboring basis functions ϕ_{i+1} and ϕ_{i-1}. This is reflected into the sparse

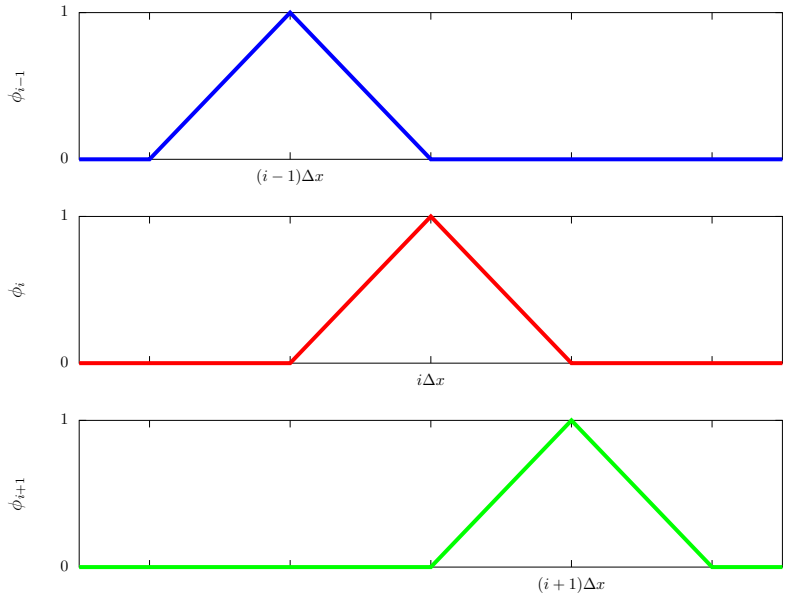


Figure A.1: Illustration of the so-called hat functions used in the linear FE spatial discretization. The basis functions $\phi_{i-1}(x)$, $\phi_i(x)$, and $\phi_{i+1}(x)$ are shown.

structure of the matrices **M** and **D**:

$$\mathbf{M} = \frac{\Delta x}{6}\begin{bmatrix} & & & & \dots & & & & \\ \dots & 0 & 1 & 4 & 1 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & 1 & 4 & 1 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & 1 & 4 & 1 & 0 & \dots \\ & & & & \dots & & & & \end{bmatrix} \tag{A.4}$$

$$\mathbf{D} = \frac{1}{4}\begin{bmatrix} & & & & \dots & & & & \\ \dots & 0 & -1 & 0 & 1 & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & -1 & 0 & 1 & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & -1 & 0 & 1 & 0 & \dots \\ & & & & \dots & & & & \end{bmatrix} \tag{A.5}$$

where Δx is the grid spacing. The values in the matrices are verified easily, by evaluating some basic integral expressions as, for example:

$$M_{i(i+1)} = M_{(i-1)i} = \int_0^{\Delta x}\left(1 - \frac{x}{\Delta x}\right)\frac{x}{\Delta x}dx = \frac{\Delta x}{6} \;\; ; \forall j = 2, ..., N-1. \tag{A.6}$$

In the previous matrices only three rows are depicted, because this is sufficient to understand the structure of the matrices. Depending on the boundary conditions that are used, the first and last row of the matrices may deviate from this pattern.
Remark that the second order finite-difference discretization Eq. (1.29) of the advection equation results in a matrix equation with the same **D** but with the identity matrix **1** as mass matrix **M**. The linear FE calculation of the first order derivative is identical to the calculation with the so-called 4^{th} order compact finite-differences or Numerov scheme, which is used for example in [123].
An important advantage of the linear FE method is that there is no transformation needed to go from grid point to spectral coefficient space. It can be seen in Figure A.1 that the coefficient U_i is, indeed, equal to the value of the field u in the grid point x_i.
By using more extended basis functions, one can construct higher-order accurate approximations. On the other hand, one will obtain less sparse matrices and more complex transformations between grid point and spectral coefficient space.

A similar analysis could be repeated in 2D. The most simple 2D basis functions are the bilinear functions $\phi_{ij}(x,y) = \psi_i(x)\,\chi_j(y)$, which consist of a product of linear basis functions. The corresponding mass and derivative matrices **M** and **D** will contain 9 and 6 elements, respectively, on each row. If we define ϕ_{ij} as the 2D chapeau function centered at (x_i, y_j) and

$$M_{ij,kl} = \int\int_\Omega \phi_{ij}(x,y)\,\phi_{kl}(x,y)\,dxdy \text{ and} \tag{A.7}$$

$$D_{ij,kl} = \int\int_\Omega \phi_{ij}(x,y)\,\frac{\partial \phi_{kl}}{\partial x}(x,y)\,dxdy, \tag{A.8}$$

then the non-zero elements of **M** and **D** are:

$$\begin{aligned}
M_{ij,ij} &= \frac{4(\Delta x)^2}{9} \\
M_{ij,(i+1)j} = M_{ij,(i-1)j} = M_{ij,i(j+1)} = M_{ij,i(j-1)} &= \frac{(\Delta x)^2}{9} \\
M_{ij,(i+1)(j+1)} = M_{ij,(i-1)(j+1)} = M_{ij,(i+1)(j-1)} = M_{ij,(i-1)(j-1)} &= \frac{(\Delta x)^2}{36} \\
D_{ij,(i+1)j} &= \frac{\Delta x}{3} \\
D_{ij,(i-1)j} &= -\frac{\Delta x}{3} \\
D_{ij,(i+1)(j+1)} = D_{ij,(i+1)(j-1)} &= \frac{\Delta x}{12} \\
D_{ij,(i-1)(j+1)} = D_{ij,(i-1)(j-1)} &= -\frac{\Delta x}{12}.
\end{aligned}$$

The previous expressions are only valid if the resolution along the x- and y-direction is identical: $\Delta x = \Delta y$.

2TL SI Z-grid scheme with implicit treatment of orography

B

This appendix shows how a local discretization method could be used to include horizontally inhomogeneous terms (in this case $\Phi_{oro}D$) in the implicit part of the calculations. The methodology is derived for the 1D SWE.
2TL SISL time discretization of the 1D continuous SWE including orography results in

$$u_A^{+(i)} + \frac{\Delta t}{2}\left(\frac{\partial \phi}{\partial x}\right)_A^{+(i)} - \frac{f\Delta t}{2} v_A^{+(i)} = R_u^{(i)}, \tag{B.1}$$

$$v_A^{+(i)} + \frac{f\Delta t}{2} u_A^{+(i)} = R_v^{(i)}, \text{ and} \tag{B.2}$$

$$\phi_A^{+(i)} + \frac{\Delta t}{2}\left(\Phi - \Phi_{oro,A}\right) D_A^{+(i)} = R_\phi^{(i)}, \quad \forall i \geqslant 0 \tag{B.3}$$

where i is the iteration index of the ICI scheme, Φ_{oro} can depend on the position, and

$$R_u^{(i)} = R_u^{(0)} = u_{D^{(i)}}^0 - \frac{\Delta t}{2}\left(\frac{\partial \phi}{\partial x}\right)_{D^{(i)}}^0 + \frac{f\Delta t}{2} v_{D^{(i)}}^0 \tag{B.4}$$

$$R_v^{(i)} = R_v^{(0)} = v_{D^{(i)}}^0 - \frac{f\Delta t}{2} u_{D^{(i)}}^0 \tag{B.5}$$

$$\begin{aligned} R_\phi^{(i)} &= \phi_{D^{(i)}}^0 - \frac{\Delta t}{2}\left(\Phi - \Phi_{oro,D^{(i)}}\right) D_{D^{(i)}}^0 - \frac{\Delta t}{2}\phi_{D^{(i)}}^0 D_{D^{(i)}}^0 \\ &- \Phi_{oro,D^{(i)}} - \frac{\Delta t}{2}\phi^{+(i-1)} D^{+(i-1)} + \Phi_{oro,A}, \quad \forall i \geqslant 1 \end{aligned} \tag{B.6}$$

$$\begin{aligned} R_\phi^{(0)} &= \phi_{D^{(i)}}^0 - \frac{\Delta t}{2}\left(\Phi - \Phi_{oro,D^{(i)}}\right) D_{D^{(i)}}^0 - \frac{\Delta t}{2}\phi_{D^{(i)}}^0 D_{D^{(i)}}^0 \\ &- \Phi_{oro,D^{(i)}} - \frac{\Delta t}{2}\phi_A^0 D_A^0 + \Phi_{oro,A}. \end{aligned}$$

In the previous expression the $\Phi_{oro}D$ term is treated in an implicit way. Reformulating Eq. (B.1) and Eq. (B.2) into equations for divergence and vorticity and introduction of

the spatial discretization yield:

$$\mathcal{P}D_A^{+(i)} + \frac{\Delta t}{2}\mathcal{P}_{xx}\phi_A^{+(i)} - \frac{f\Delta t}{2}\mathcal{P}Z_A^{+(i)} = \mathcal{P}_x R_u^{(i)} \text{ and} \tag{B.7}$$

$$\mathcal{P}Z_A^{+(i)} + \frac{f\Delta t}{2}\mathcal{P}D_A^{+(i)} = \mathcal{P}_x R_v^{(i)}. \tag{B.8}$$

Elimination of vorticity $Z_A^{+(i)}$ in Eq. (B.7) and Eq. (B.8) then yields:

$$X\mathcal{P}D_A^{+(i)} + \frac{\Delta t}{2}\mathcal{P}_{xx}\phi_A^{+(i)} = \mathcal{P}_x R_u^{(i)} + \frac{f\Delta t}{2}\mathcal{P}_x R_v^{(i)} \quad \text{with} \quad X = 1 + \left(\frac{f\Delta t}{2}\right)^2. \tag{B.9}$$

Dividing Eq. (B.3) by $\Phi - \Phi_{oro,A}$ and doing then the spatial discretization results in:

$$\mathcal{P}^*\phi_A^{+(i)} + \frac{\Delta t}{2}\mathcal{P}D_A^{+(i)} = \mathcal{P}^* R_\phi^{(i)}. \tag{B.10}$$

The orography term $\frac{1}{\Phi-\Phi_{oro}}$ is included in the operator $\mathcal{P}^*$. Let us illustrate this for the situation when a linear FE method is used. The mass matrix $\mathbf{M}$ given by Eq. (A.4) and corresponding to the operator $\mathcal{P}$, is then replaced by:

$$\mathbf{M}^* = \frac{\Delta x}{6}\begin{bmatrix} & & & & \dots & & & \\ \dots & 0 & \frac{1}{\Phi-\Phi_{oro}(x_{i-2})} & \frac{4}{\Phi-\Phi_{oro}(x_{i-1})} & \frac{1}{\Phi-\Phi_{oro}(x_i)} & 0 & 0 & 0 & \dots \\ \dots & 0 & 0 & \frac{1}{\Phi-\Phi_{oro}(x_{i-1})} & \frac{4}{\Phi-\Phi_{oro}(x_i)} & \frac{1}{\Phi-\Phi_{oro}(x_{i+1})} & 0 & 0 & \dots \\ \dots & 0 & 0 & 0 & \frac{1}{\Phi-\Phi_{oro}(x_i)} & \frac{4}{\Phi-\Phi_{oro}(x_{i+1})} & \frac{1}{\Phi-\Phi_{oro}(x_{i+2})} & 0 & \dots \\ & & & & \dots & & & & \end{bmatrix}, \tag{B.11}$$

with i the column index of the matrix. Including the horizontally varying orography, breaks the symmetry of the matrix and this can have an effect on the efficiency of the solver. Combining Eq. (B.9) and Eq. (B.10) results in a Helmholtz equation for geopotential:

$$\left[X\mathcal{P}^* - \left(\frac{\Delta t}{2}\right)^2 \mathcal{P}_{xx}\right]\phi_A^{+(i)} = X\mathcal{P}^* R_\phi^{(i)} - \frac{\Delta t}{2}\mathcal{P}_x R_u^{(i)} - f\left(\frac{\Delta t}{2}\right)^2 \mathcal{P}_x R_v^{(i)}. \tag{B.12}$$

Equations (B.12), (B.10), and (B.8) permit to update geopotential, divergence, and vorticity.

One can also construct a Eulerian version of a scheme treating orography implicitly. A careful analysis shows that the scheme will be identical except for the right hand

sides. For the first iteration ($i = 0$) we have

$$R_u^{(0)} = u^0 - \frac{\Delta t}{2}\left(\frac{\partial \phi}{\partial x}\right)^0 + \frac{f\Delta t}{2}v^0 - \Delta t u^0 D^0, \tag{B.13}$$

$$R_v^{(0)} = v^0 - \frac{f\Delta t}{2}u^0 - \Delta t u^0 Z^0 \text{ , and} \tag{B.14}$$

$$R_\phi^{(0)} = \phi^0 - \Delta t u^0 \left(\frac{\partial \phi}{\partial x}\right)^0 - \frac{\Delta t}{2}\left(\Phi - \Phi_{oro}\right) D^0 + \Delta t u^0 \frac{\partial \phi_{oro}}{\partial x} - \Delta t \phi^0 D^0 \tag{B.15}$$

and for the following iterations ($i \geqslant 1$)

$$R_u^{(i)} = u^0 - \frac{\Delta t}{2}\left(\frac{\partial \phi}{\partial x}\right)^0 + \frac{f\Delta t}{2}v^0 - \frac{\Delta t}{2}u^0 D^0 - \frac{\Delta t}{2}u^{+(i-1)} D^{+(i-1)}, \tag{B.16}$$

$$R_v^{(i)} = v^0 - \frac{f\Delta t}{2}u^0 - \frac{\Delta t}{2}u^0 Z^0 - \frac{\Delta t}{2}u^{+(i-1)} Z^{+(i-1)} \text{ , and} \tag{B.17}$$

$$\begin{aligned} R_\phi^{(i)} &= \phi^0 - \frac{\Delta t}{2}u^0 \left(\frac{\partial \phi}{\partial x}\right)^0 - \frac{\Delta t}{2}u^{+(i-1)} \left(\frac{\partial \phi}{\partial x}\right)^{+(i-1)} - \frac{\Delta t}{2}\left(\Phi - \Phi_{oro}\right) D^0 \\ &+ \frac{\Delta t}{2}\left(u^0 + u^{+(i-1)}\right)\frac{\partial \phi_{oro}}{\partial x} - \frac{\Delta t}{2}\left(\phi^0 D^0 + \phi^{+(i-1)} D^{+(i-1)}\right). \end{aligned} \tag{B.18}$$

Routine to calculate responses for FD A-grid and Z-grid scheme

C

There exists both an A-grid and Z-grid version of the routine that calculates the FD responses.

C.1 A-grid version

```
MODULE ERESP_MOD

CONTAINS

REAL(KIND=JPRB) FUNCTION FRESPL(PKL)

! Definition function responses d/dx
USE PARKIND1  ,ONLY : JPIM     ,JPRB
USE YOMHOOK   ,ONLY : LHOOK,   DR_HOOK

USE YOMZFD    ,ONLY : LDISLOC, NORDER
USE YOMLUN    ,ONLY : NULOUT
USE YEMGEO    ,ONLY : EDELX
USE TPM_ALD_GEO, ONLY : GALD

IMPLICIT NONE

INTEGER(KIND=JPIM), INTENT(IN) :: PKL

REAL(KIND=JPRB) :: ZHOOK_HANDLE
REAL(KIND=JPRB) :: ARG

IF (LHOOK) CALL DR_HOOK('FRESPL',0,ZHOOK_HANDLE)

FRESPL = 0.0_JPRB
ARG = REAL(PKL,JPRB)*GALD%EXWN*EDELX

IF (LDISLOC) THEN
  IF (NORDER.EQ.2) THEN
    FRESPL = SIN(ARG)/EDELX
  ELSEIF (NORDER.EQ.4) THEN
    FRESPL = REAL(4,JPRB)/REAL(3,JPRB)*SIN(ARG)
    FRESPL = FRESPL - REAL(1,JPRB)/REAL(6,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPL = FRESPL/EDELX
  ELSEIF (NORDER.EQ.6) THEN
    FRESPL = REAL(3,JPRB)/REAL(2,JPRB)*SIN(ARG)
    FRESPL = FRESPL - REAL(3,JPRB)/REAL(10,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPL = FRESPL + REAL(1,JPRB)/REAL(30,JPRB)*SIN(REAL(3,JPRB)*ARG)
    FRESPL = FRESPL/EDELX
  ELSEIF (NORDER.EQ.8) THEN
    FRESPL = REAL(8,JPRB)/REAL(5,JPRB)*SIN(ARG)
    FRESPL = FRESPL - REAL(2,JPRB)/REAL(5,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPL = FRESPL + REAL(8,JPRB)/REAL(105,JPRB)*SIN(REAL(3,JPRB)*ARG)
    FRESPL = FRESPL - REAL(1,JPRB)/REAL(140,JPRB)*SIN(REAL(4,JPRB)*ARG)
    FRESPL = FRESPL/EDELX
  ELSE
    FRESPL = REAL(0,JPRB)
  ENDIF
ELSE
  FRESPL = ARG/EDELX
```

```
51 ENDIFIF (LHOOK) CALL DR_HOOK('FRESPL',1,ZHOOK_HANDLE)
52
53 END FUNCTION FRESPL
54 !#####################################################################
55 REAL(KIND=JPRB) FUNCTION FRESPM(PKM)
56
57 ! Definition function responses d/dy
58 USE PARKIND1  ,ONLY : JPIM     ,JPRB
59 USE YOMHOOK   ,ONLY : LHOOK,   DR_HOOK
60
61 USE YOMZFD    ,ONLY : LDISLOC, NORDER
62 USE YOMLUN    ,ONLY : NULOUT
63 USE YEMGEO    ,ONLY : EDELY
64 USE TPMALD_GEO, ONLY : GALD
65
66 IMPLICIT NONE
67
68 INTEGER(KIND=JPIM), INTENT(IN) :: PKM
69
70 REAL(KIND=JPRB) :: ZHOOK_HANDLE
71 REAL(KIND=JPRB) :: ARG
72
73 IF (LHOOK) CALL DR_HOOK('FRESPM',0,ZHOOK_HANDLE)
74
75 ARG = REAL(PKM,JPRB)*GALD%EYWN*EDELY
76
77 FRESPM = 0.0_JPRB
78
79 IF (LDISLOC) THEN
80   IF (NORDER.EQ.2) THEN
81     FRESPM = SIN(ARG)/EDELY
82   ELSEIF (NORDER.EQ.4) THEN
83     FRESPM = REAL(4,JPRB)/REAL(3,JPRB)*SIN(ARG)
84     FRESPM = FRESPM - REAL(1,JPRB)/REAL(6,JPRB)*SIN(REAL(2,JPRB)*ARG)
85     FRESPM = FRESPM/EDELY
86   ELSEIF (NORDER.EQ.6) THEN
87     FRESPM = REAL(3,JPRB)/REAL(2,JPRB)*SIN(ARG)
88     FRESPM = FRESPM - REAL(3,JPRB)/REAL(10,JPRB)*SIN(REAL(2,JPRB)*ARG)
89     FRESPM = FRESPM + REAL(1,JPRB)/REAL(30,JPRB)*SIN(REAL(3,JPRB)*ARG)
90     FRESPM = FRESPM/EDELY
91   ELSEIF (NORDER.EQ.8) THEN
92     FRESPM = REAL(8,JPRB)/REAL(5,JPRB)*SIN(ARG)
93     FRESPM = FRESPM - REAL(2,JPRB)/REAL(5,JPRB)*SIN(REAL(2,JPRB)*ARG)
94     FRESPM = FRESPM + REAL(8,JPRB)/REAL(105,JPRB)*SIN(REAL(3,JPRB)*ARG)
95     FRESPM = FRESPM - REAL(1,JPRB)/REAL(140,JPRB)*SIN(REAL(4,JPRB)*ARG)
96     FRESPM = FRESPM/EDELY
97   ELSE
98     FRESPM = REAL(0,JPRB)
99   ENDIF
100 ELSE
101   FRESPM = ARG/EDELY
102 ENDIF
103
104 IF (LHOOK) CALL DR_HOOK('FRESPM',1,ZHOOK_HANDLE)
105
106 END FUNCTION FRESPM
107 !#####################################################################
108 REAL(KIND=JPRB) FUNCTION FLAPRESP(PKL,PKM)
109
110 ! Definition function laplacian
111 USE PARKIND1  ,ONLY : JPIM     ,JPRB
112 USE YOMHOOK   ,ONLY : LHOOK,   DR_HOOK
113
114 USE YOMLUN    ,ONLY : NULOUT
115
116 IMPLICIT NONE
117
118 INTEGER(KIND=JPIM), INTENT(IN) :: PKM  ,PKL
119
120 REAL(KIND=JPRB) :: ZHOOK_HANDLE
121 IF (LHOOK) CALL DR_HOOK('FLAPRESP',0,ZHOOK_HANDLE)
122
123 FLAPRESP = 0.0_JPRB
124 FLAPRESP = FRESPM(PKM)**2 +  FRESPL(PKL)**2
125
126 IF (LHOOK) CALL DR_HOOK('FLAPRESP',1,ZHOOK_HANDLE)
127
128 END FUNCTION FLAPRESP
129 END MODULE ERESP_MOD
```

C.2 Z-grid version

```
MODULE ERESP_MOD

CONTAINS

REAL(KIND=JPRB) FUNCTION FRESPL(PKL)

! Definition function responses d/dx
USE PARKIND1   ,ONLY : JPIM        ,JPRB
USE YOMHOOK    ,ONLY : LHOOK,    DR_HOOK

USE YOMZFD     ,ONLY : LDISLOC, NORDER
USE YOMLUN     ,ONLY : NULOUT
USE YEMGEO     ,ONLY : EDELX
USE TPMALD_GEO ,ONLY: GALD

IMPLICIT NONE

INTEGER(KIND=JPIM), INTENT(IN) :: PKL

REAL(KIND=JPRB) :: ZHOOK_HANDLE
REAL(KIND=JPRB) :: ARG

IF (LHOOK) CALL DR_HOOK('FRESPL',0,ZHOOK_HANDLE)

FRESPL = 0.0_JPRB
ARG = REAL(PKL,JPRB)*GALD%EXWN*EDELX

IF (LDISLOC) THEN
  IF (NORDER.EQ.2) THEN
    FRESPL = SIN(ARG)/EDELX
  ELSEIF (NORDER.EQ.4) THEN
    FRESPL = REAL(4,JPRB)/REAL(3,JPRB)*SIN(ARG)
    FRESPL = FRESPL - REAL(1,JPRB)/REAL(6,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPL = FRESPL/EDELX
  ELSEIF (NORDER.EQ.6) THEN
    FRESPL = REAL(3,JPRB)/REAL(2,JPRB)*SIN(ARG)
    FRESPL = FRESPL - REAL(3,JPRB)/REAL(10,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPL = FRESPL + REAL(1,JPRB)/REAL(30,JPRB)*SIN(REAL(3,JPRB)*ARG)
    FRESPL = FRESPL/EDELX
  ELSEIF (NORDER.EQ.8) THEN
    FRESPL = REAL(8,JPRB)/REAL(5,JPRB)*SIN(ARG)
    FRESPL = FRESPL - REAL(2,JPRB)/REAL(5,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPL = FRESPL + REAL(8,JPRB)/REAL(105,JPRB)*SIN(REAL(3,JPRB)*ARG)
    FRESPL = FRESPL - REAL(1,JPRB)/REAL(140,JPRB)*SIN(REAL(4,JPRB)*ARG)
    FRESPL = FRESPL/EDELX
  ELSE
    FRESPL = REAL(0,JPRB)
  ENDIF
ELSE
  FRESPL = ARG/EDELX
ENDIF
IF (LHOOK) CALL DR_HOOK('FRESPL',1,ZHOOK_HANDLE)

END FUNCTION FRESPL
!#####################################################################
REAL(KIND=JPRB) FUNCTION FRESPM(PKM)

! Definition function responses d/dy
USE PARKIND1   ,ONLY : JPIM        ,JPRB
USE YOMHOOK    ,ONLY : LHOOK,    DR_HOOK

USE YOMZFD     ,ONLY : LDISLOC, NORDER
USE YOMLUN     ,ONLY : NULOUT
!USE YEMGEO     , ONLY : EYWN,   EDELY
USE YEMGEO     ,ONLY : EDELY
USE TPMALD_GEO ,ONLY: GALD

IMPLICIT NONE

INTEGER(KIND=JPIM), INTENT(IN) :: PKM

REAL(KIND=JPRB) :: ZHOOK_HANDLE
REAL(KIND=JPRB) :: ARG

```

```
IF (LHOOK) CALL DR_HOOK('FRESPM',0,ZHOOK_HANDLE)

ARG = REAL(PKM,JPRB)*GALD%EYWN*EDELY

FRESPM = 0.0_JPRB

IF (LDISLOC) THEN
  IF (NORDER.EQ.2) THEN
    FRESPM = SIN(ARG)/EDELY
  ELSEIF (NORDER.EQ.4) THEN
    FRESPM = REAL(4,JPRB)/REAL(3,JPRB)*SIN(ARG)
    FRESPM = FRESPM - REAL(1,JPRB)/REAL(6,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPM = FRESPM/EDELY
  ELSEIF (NORDER.EQ.6) THEN
    FRESPM = REAL(3,JPRB)/REAL(2,JPRB)*SIN(ARG)
    FRESPM = FRESPM - REAL(3,JPRB)/REAL(10,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPM = FRESPM + REAL(1,JPRB)/REAL(30,JPRB)*SIN(REAL(3,JPRB)*ARG)
    FRESPM = FRESPM/EDELY
  ELSEIF (NORDER.EQ.8) THEN
    FRESPM = REAL(8,JPRB)/REAL(5,JPRB)*SIN(ARG)
    FRESPM = FRESPM - REAL(2,JPRB)/REAL(5,JPRB)*SIN(REAL(2,JPRB)*ARG)
    FRESPM = FRESPM + REAL(8,JPRB)/REAL(105,JPRB)*SIN(REAL(3,JPRB)*ARG)
    FRESPM = FRESPM - REAL(1,JPRB)/REAL(140,JPRB)*SIN(REAL(4,JPRB)*ARG)
    FRESPM = FRESPM/EDELY
  ELSE
    FRESPM = REAL(0,JPRB)
  ENDIF
ELSE
  FRESPM = ARG/EDELY
ENDIF

IF (LHOOK) CALL DR_HOOK('FRESPM',1,ZHOOK_HANDLE)

END FUNCTION FRESPM
!#####################################################################
REAL(KIND=JPRB) FUNCTION FLAPRESP(PKL,PKM)

! Definition function laplacian
USE PARKIND1  ,ONLY : JPIM     ,JPRB
USE YOMHOOK   ,ONLY : LHOOK,   DR_HOOK

USE YOMZFD    ,ONLY : LDISLOC, NORDER
USE YOMLUN    ,ONLY : NULOUT
!USE YEMGEO    , ONLY : EXWN     ,EYWN   ,EDELX    ,EDELY
USE YEMGEO    ,ONLY : EDELX    ,EDELY
USE TPM_ALD_GEO ,ONLY: GALD

IMPLICIT NONE

INTEGER(KIND=JPIM), INTENT(IN) :: PKM   ,PKL

REAL(KIND=JPRB) :: ZHOOK_HANDLE
REAL(KIND=JPRB) :: REPL, REPM
REAL(KIND=JPRB) :: ARGX, ARGY
IF (LHOOK) CALL DR_HOOK('FLAPRESP',0,ZHOOK_HANDLE)

FLAPRESP = 0.0_JPRB
ARGX = REAL(PKL,JPRB)*GALD%EXWN*EDELX
ARGY = REAL(PKM,JPRB)*GALD%EYWN*EDELY

IF (LDISLOC) THEN
  IF (NORDER.EQ.2) THEN
    FLAPRESP = (REAL(2,JPRB) - REAL(2,JPRB)*COS(ARGX))/(EDELX*EDELX)
    FLAPRESP = FLAPRESP + (REAL(2,JPRB) - REAL(2,JPRB)*COS(ARGY))/(EDELY*EDELY)
  ELSEIF (NORDER.EQ.4) THEN
    FLAPRESP = (REAL(5,JPRB)/REAL(2,JPRB) - REAL(8,JPRB)/REAL(3,JPRB)*COS(ARGX))/EDELX**2
    FLAPRESP = FLAPRESP + REAL(1,JPRB)/REAL(6,JPRB)*COS(REAL(2,JPRB)*ARGX)/EDELX**2
    FLAPRESP = FLAPRESP + (REAL(5,JPRB)/REAL(2,JPRB) - REAL(8,JPRB)/REAL(3,JPRB)*COS(ARGY))/EDELY**2
    FLAPRESP = FLAPRESP + REAL(1,JPRB)/REAL(6,JPRB)*COS(REAL(2,JPRB)*ARGY)/EDELY**2
  ELSEIF (NORDER.EQ.6) THEN
    FLAPRESP = (REAL(49,JPRB)/REAL(18,JPRB) - REAL(3,JPRB)*COS(ARGX))/EDELX**2
    FLAPRESP = FLAPRESP + REAL(3,JPRB)/REAL(10,JPRB)*COS(REAL(2,JPRB)*ARGX)/EDELX**2
    FLAPRESP = FLAPRESP - REAL(1,JPRB)/REAL(45,JPRB)*COS(REAL(3,JPRB)*ARGX)/EDELX**2
    FLAPRESP = FLAPRESP + (REAL(49,JPRB)/REAL(18,JPRB) - REAL(3,JPRB)*COS(ARGY))/EDELY**2
    FLAPRESP = FLAPRESP + REAL(3,JPRB)/REAL(10,JPRB)*COS(REAL(2,JPRB)*ARGY)/EDELY**2
    FLAPRESP = FLAPRESP - REAL(1,JPRB)/REAL(45,JPRB)*COS(REAL(3,JPRB)*ARGY)/EDELY**2
  ELSEIF (NORDER.EQ.8) THEN
    FLAPRESP = (REAL(205,JPRB)/REAL(72,JPRB) - REAL(16,JPRB)/REAL(5,JPRB)*COS(ARGX))/EDELX**2
    FLAPRESP = FLAPRESP + REAL(2,JPRB)/REAL(5,JPRB)*COS(REAL(2,JPRB)*ARGX)/EDELX**2
```

```
      FLAPRESP = FLAPRESP - REAL(16,JPRB)/REAL(315,JPRB)*COS(REAL(3,JPRB)*ARGX)/EDELX**2
      FLAPRESP = FLAPRESP + REAL(1,JPRB)/REAL(280,JPRB)*COS(REAL(4,JPRB)*ARGX)/EDELX**2
      FLAPRESP = FLAPRESP + (REAL(205,JPRB)/REAL(72,JPRB) - REAL(16,JPRB)/REAL(5,JPRB)*COS(ARGY))/EDELY**2
      FLAPRESP = FLAPRESP + REAL(2,JPRB)/REAL(5,JPRB)*COS(REAL(2,JPRB)*ARGY)/EDELY**2
      FLAPRESP = FLAPRESP - REAL(16,JPRB)/REAL(315,JPRB)*COS(REAL(3,JPRB)*ARGY)/EDELY**2
      FLAPRESP = FLAPRESP + REAL(1,JPRB)/REAL(280,JPRB)*COS(REAL(4,JPRB)*ARGY)/EDELY**2
    ELSE
      FLAPRESP = REAL(0,JPRB)
    ENDIF
  ELSE
    FLAPRESP = (ARGX/EDELX)**2 + (ARGY/EDELY)**2
  ENDIF

  IF (LHOOK) CALL DR_HOOK('FLAPRESP',1,ZHOOK_HANDLE)

END FUNCTION FLAPRESP
END MODULE ERESP_MOD
```

Bibliography

[1] Abdalla S., Isaksen L., Janssen PAEM, Wedi N.: Effective spectral resolution of IFS. ECMWF Newsletter. **137**, 19–22 (2013)

[2] ALADIN International Team: The ALADIN project: mesoscale modelling seen as a basic tool for weather forecasting and atmospheric research. WMO Bull. **46**, 317–324 (1997)

[3] Design of the UCLA general circulation model. Technical report 7. Department of Meteorology, University of California (1972)

[4] Arakawa A.: Adjustment mechanisms in atmospheric models. J. Meteorol. Soc. Jpn. **75B**, 155–179 (1997)

[5] Arakawa A., Konor C.: Unification of the Anelastic and Quasi-Hydrostatic Systems of Equations. Mon. Wea. Rev. **137**, 710–726 (2009)

[6] Bénard P., Mašek J, Smolíková P.: Stability of leapfrog constant coefficients semi-implicit schemes for the fully elastic system of Euler equations: Case with orography. Mon. Wea. Rev. **133**, 1065–1074 (2005)

[7] Bénard P., Mašek J.: Scientific documentation for ALADIN NH dynamical kernel (Version 3.1.0). can be consulted at http://www.cnrm.meteo.fr/gmapdoc//IMG/pdf/designv3_1_0.pdf (2013)

[8] Bénard P.: An assessment of global forecast errors due to the spherical geopotential approximation in the shallow-water case. Q. J. R. Meteorol. Soc. **141**, 195–206 (2014)

[9] Bijlsma S.J.: Optimal Aircraft Routing in General Wind Fields. J. of Guid., Contr., and Dyn. **32**, 1025–1029 (2009)

[10] Bjerkness V.: Das problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Phyisik. Meteorol. Z. **21**, 1–7 (1904)

[11] Boyd J.P.: Chebyshev and Fourier Spectral Methods. Dover, pp 688 (2001)

[12] Boutter F., Courtier P.: Data assimilation concepts and methods. ECMWF lecture notes, pp 59 (1999)

[13] Bourke W.: An efficient, one-level, primitive equation spectral model. Mon. Wea. Rev. **100**, 683–689 (1972)

[14] Brown A., Milton S., Cullen M., Golding B., Mitchell J., Shelly A.: Unified modeling and prediction of weather and climate. A 25-Year Journey. Bull. Amer. Meteor. Soc. **93**, 1865–1877 (2012)

[15] Caluwaerts S., Degrauwe D., Termonia P., Voitus F., Bénard P., Geleyn J.F.: Importance of temporal symmetry in spatial discretization for geostrophic adjustment in semi-implicit Z-grid schemes. Q. J. R. Meteorol. Soc. **141**, 128–138 (2015)

[16] Caluwaerts S., Degrauwe D., Voitus F., Termonia P.: Disretization in Numerical Weather Prediction: a modular approach to investigate spectral and local SISL methods. ECMI Special Volume 'Mathematical Problems in Meteorological Modelling', accepted for publication

[17] Cats G.: 24 more years of Numerical Weather Prediction: A model performance model. Unpublished report, pp 31 (2008)

[18] Charney J.G., Fjörtoft R., Von Neumann J.: Numerical integration of the barotropic vorticity equation. Tellus. **2**, 237–254 (1950)

[19] Coiffier J.: Fundamentals of Numerical Weather Prediction. Cambridge University Press, pp 340 (2011)

[20] Côté J., Staniforth A.: An accurate and efficient finite-element global model of the shallow water equations. Mon. Weather Rev. **118**, 2707–2717 (1990)

[21] Côté J., Roch M., Staniforth A., Fillion L.: A variable-resolution semi-Lagrangian finite-element global model of the shallow-water equations. Mon. Weather Rev. **121**, 231–243 (1993)

[22] Côté J., Gravel S., Méthot A., Patoine A., Roch M., Staniforth A.: The operational CMC-MRB global environmental multiscale (GEM) model. Part 1: Design considerations and formulation. Mon. Weather Rev. **126**, 1373–1395 (1998)

[23] Cotter CJ., Shipton J.: Mixed finite elements for numerical weather prediction. J. Comput. Phys. **231**, 7076–7091 (2012)

[24] Courant R., Friedrichs K.O., Lewy H..: Uber die Partielen Differenzengleichungen der Mathematischen Physik. Math. Annalen. **100**, 32–74 (1928)

[25] Courtier P., Geleyn J.F.: A global numerical weather prediction model with variable resolution: application to the shallow water equations. **114**, 1321–1346 (1988)

[26] Davies H.: A lateral boundary formulation for multi-level prediction models. Q. J. R. Meteorol. Soc. **102**, 405–418 (1976)

[27] Davies T., Staniforth A., Wood N., Thuburn J.: Validity of anelastic and other equation sets as inferred from normal-mode analysis. Q. J. R. Meteorol. Soc. **129**, 2761–2775 (2003)

[28] Davies T., Cullen M., Malcolm A., Mawson M., Staniforth A., White A., Wood N.: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. Q. J. R. Meteorol. Soc. **131**, 1759–1782 (2005)

[29] Davies T.: Lateral boundary conditions for limited area models. Q. J. R. Meteorol. Soc. **140**, 185–196 (2014)

[30] Degrauwe D., Caluwaerts S., Voitus F., Hamdi R., Termonia P.: Application of Boyd's periodization and relaxation method in a spectral Atmospheric Limited-Area Model. Part 2: Accuracy analysis and detailed study of operational impact. Mon. Weather Rev. **140**, 3149–3162 (2012)

[31] Dennis J., Edwards J., Evans K., Guba O., Lauritzen P., Mirin A., St-Cyr A., Taylor M., Worley P.: CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. Int. J. of High Perform. Appl. **26**, 74–89 (2011)

[32] De Troch R., Hamdi R., Van de Vyver H., Geleyn J.F., Termonia P.: Multiscale performance of the ALARO-0 model for simulating extreme summer precipitation climatology in Belgium. J. of Clim. **26**, 8895–8915 (2013)

[33] Dubos T., Voitus F.: A Semihydrostatic Theory of Gravity-Dominated Compressible Flow. J. Atmos. Sci. **71**, 4621–4638 (2014)

[34] Durran D.R.: Numerical Methods for Fluid Dynamics with applications to Geophysics. Springer, pp 516 (2010)

[35] Durran D., Blossey P.: Implicit-explicit multistep methods for fast-wave-slow-wave problems. Mon. Weather Rev. **140**, 1307–1325 (2012)

[36] Eckart C.: The hydrodynamics of oceans and atmospheres. Pergamon Press, pp 290 (1960)

[37] ECMWF: IFS documentation: Cy40r1 operational implemantation 22 november 2013. PART VI: Technical and computational procedures. can be consulted at http://old.ecmwf.int/research/ifsdocs/CY40r1/IFSPart6.pdf (2015)

[38] IPCC: Climate Change 2014: Working group 2 contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. can be consulted at http://www.ipcc.ch/report/ar5/ (2014)

[39] Eliasen E., Machenhauer B., Rasmussen E.: On a numerical method for integration of the hydrodynamical equations with a spectral representation of the hori-

zontal fields. Report nr 2. Institut for Teoretisk Meteorologi, University of Copenhagen (1970)

[40] Walters D., Wood N., Vosper S., Milton S.: ENDGame: A new dynamical core for seamless atmospheric prediction. Met Office documentation. can be consulted at http://www.metoffice.gov.uk/media/pdf/s/h/ENDGameGOVSci_v2.0.pdf (2014)

[41] Fjortoft R.: On a numerical method of integrating the barotropic vorticity equation. Tellus. **4**, 179–194 (1952)

[42] Fox-Rabinovitz M., Côté J., Dugas B., Déqué M., McGregor J.: Variable resolution general circulation models: stretched-grid model intercomparison project (SGMIP). J. Geophys. Res. **111** (2006)

[43] Fritz W.B.: ENIAC - A problem solver. IEEE Ann. Hist. Comp. **16**, 25–45 (1994)

[44] Gassmann A., Herzog H.: A consistent time-split numerical scheme applied to the nonhydrostatic compressible equations. Mon. Weather Rev. **135**, 20–36 (2006)

[45] Gerard L.: An integrated package for subgrid convection, clouds and precipitation compatible with meso-gamma scales. Q. J. R. Meteorol. Soc. **133**, 711–730 (2007)

[46] Giot O., Termonia P., Degrauwe D., De Troch R., Caluwaerts S., Smet G., Berckmans J., Deckmyn A., De Cruz L., De Meutter P., Duerinckx A., Gerard L., Hamdi R., Van den Bergh J., Van Ginderachter M., Van Schaeybroeck B.: Validation of the ALARO-0 model within the EURO-CORDEX framework, Geosci. Model Dev. Discuss., **8**, 8387–8409 (2015)

[47] Godunov S.K.: A difference scheme for numerical computation of discontinuous solutions of equations in fluid dynamics. Mat. Sb. **47**, 271 (1959)

[48] Golnaraghi M., Etienne C., Sapir D.G., Below R.: Atlas of mortality and economic losses from weather, climate and water extremes (1970–2012). WMO report 1123, 1–46 (2014)

[49] Hawkins M., Weger I.: Supercomputing at ECMWF. ECMWF Newsletter. **143**, 32–38 (2015)

[50] Holton J.R.: An introduction to dynamic meteorology. Elsevier Academic Press, pp 535 (2004)

[51] Honnert R., Masson V.: What is the smallest physically acceptable scale for 1D-turbulence schemes? Front. Earth Sci. **2**, 1–5 (2014)

[52] Hortal M.: The development and testing of a new two-time level semi-Lagrangian scheme (SETTLS) in the ECMWF forecast model. Q. J. R. Meteorol. Soc. **128**, 1671–1687 (2002)

[53] Hortal M., Simmons A.J.: Use of reduced Gaussian grids in spectral models. Mon. Weather Rev. **119**, 1057–1074 (1991)

[54] Hurrell J., Meehl G., Bader D., Delworth T., Kirtman B., Wielicki B.: A Unified Modeling Approach to Climate System Prediction. Bull. Amer. Meteor. Soc. **90**, 1819–1832 (2009)

[55] Ikawa M.: Comparison of some schemes for nonhydrostatic models with orography. J. Meteorol. Soc. Jpn. **66**, 1065–1075 (1988)

[56] JAM: Outline of the operational numerical weather prediction at the Japan Meteorological Agency: http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline-nwp/index.htm (2015)

[57] Klemp J., Wilhelmson R.: The simulation of three-dimensional convective storm dynamics. J. Atmos. Sci. **35**, 1070–1096 (1978)

[58] Kwizak M., Robert A.: A semi-implicit scheme for grid point atmospheric models of the primitive equations. Mon. Weather. Rev. **99**, 32–36 (1971)

[59] Laprise R.: The Euler equations of motion with hydrostatic pressure as an independent variable. Mon. Weather. Rev. **120**, 197–207 (1992)

[60] Laprise R.: Notes and Correspondence: Resolved scales and nonlinear interactions in Limited-Area Models. J. Atmos. Sci. **60**, 768–779 (2002)

[61] Lin S.: A vertically Lagrangian Finite-Volume dynamical core for global models. Mon. Weather. Rev. **132**, 2293–2307 (2004)

[62] Lynch P., Huang X.-Y.: Initialization of the HIRLAM model using a digital filter. Mon. Weather. Rev. **120**,1019–1034 (1992)

[63] Lynch P.: The emergence of Numerical Weather Prediction: Richardson's dream. Cambridge University Press, pp 272 (2006)

[64] Majewski D., Liermann D., Prohl P., Ritter B., Buchhold M., Hanisch T., Paul G., Wergen W., Baumgardner J.: The operational global isocahedral-hexagonal gridpoint model GME: description and high-resolution tests. Mon. Weather. Rev. **130**, 319–338 (2002)

[65] Malardel S., Wedi N., Deconinck W., Diamantakis M., Kühnlein C., Mozdzynski G., Hamrud M., Smolarkiewicz P.: A new grid for the IFS. **146**, 23–28 (2016)

[66] Markowski P., Richardson Y.: Mesoscale Meteorology in Midlatitudes. Wiley-Blackwell, pp 430 (2010)

[67] Melton R., Wills L.: An analysis of the spectral transform operations in climate and weather models. SIAM J. Sci. Comput. **31**, 167–188 (2008)

[68] Mesinger F., Arakawa A.: Numerical methods used in atmospheric models. GARP Pub. Series No. 17 1. World Meteorological Organisation: Geneva (1976)

[69] Müller E., Scheichl R.: Massively parallel solvers for elliptic partial differential equations in numerical weather and climate prediction. Q. J. R. Meteorol. Soc. **140**, 2605–2624 (2014)

[70] Melvin T., Staniforth A., Thuburn J.: Dispersion analysis of the spectral element method. Q. J. R. Meteorol. Soc. **138**, 1934–1947 (2012)

[71] Melvin T., Staniforth A., Cotter C.: A two-dimensional mixed finite-element pair on rectangles. Q. J. R. Meteorol. Soc. **140**, 930–942 (2013)

[72] ALADIN: Fourth Memorandum of Understanding for the ALADIN consortium. http://www.cnrm.meteo.fr/aladin/IMG/pdf/MoU-4_def.pdf (2010)

[73] Michioka T., Sato A., Sada K.: Large-eddy simulations coupled to mesoscale meteorological model for gas dispersion in an urban district. Atm. Env. **75**, 153–162 (2013)

[74] Miller M.: Meteorological training course lecture series ECMWF: Atmospheric Waves. http://www.ecmwf.int/sites/default/files/Atmospheric

[75] Nair D., Thomas S., Loft R.: A discontinuous Galerkin global shallow water model. Mon. Weather. Rev. **133**, 876–888 (2005)

[76] Neta B.,Williams R.: Rossby Wave Frequencies and Group Velocities for Finite Element and Finite Difference Approximations to the Vorticity-Divergence and the Primitive Forms of the Shallow Water Equations. Mon. Wea. Rev. **117**, 1439–1457 (1989)

[77] Notay Y., Napov A.: A massively parallel solver for discrete Poisson-like problems. J. Comput. Phys. **281**, 237–250 (2015)

[78] Omrani H., Drobinski P., Dubos T.: Investigation of indiscriminate nudging and predictability in a nested quasi-geostrophic model. Q. J. R. Meteorol. Soc. **138**, 158–169 (2012)

[79] Orszag SA.: Transform method for calculation of vector-coupled sums: application to the spectral form of the vorticity equation. J. Atmos. Sci. **27**, 890–895 (1970)

[80] Persson A.: Notes and correspondence: Is the Coriolis effect an 'optical illusion'? Q. J. R. Meteorol. Soc. **141**, 1957–1967 (2015)

[81] Phillips N.: A coordinate system having some special advantages for numerical forecasting. J. Meteor. **14**, 184–185 (1957)

[82] Purser R.J., Leslie L.M.: A semi-implicit semi-Lagrangian finite-difference scheme using high-order spatial differencing on a nonstaggered grid. Mon. Wea. Rev. **116**, 2069–2080 (1988)

[83] Qaddouri A., Lee V.: The Canadian Global Environmental Multiscale model on the Yin-Yang grid system. Q. J. R. Meteorol. Soc. **137**, 1913–1926 (2011)

[84] Rabier F.: Overview of global data assimilation developments in numerical weather production. Q. J. R. Meteorol. Soc. **131**, 3215–3233 (2005)

[85] Richardson L.F.: Weather prediction by numerical process. Cambridge, the University press, pp 236 (1922)

[86] Randall D.: Geostrophic adjustment and the finite-difference shallow water equations. Mon. Weather. Rev. **122**, 1371–1377 (1994)

[87] Radnoti G.: Comments on a 'A spectral limited-area formulation with time-dependent boundary conditions applied to the shallow-water equations'. Mon. Weather. Rev. **123**, 3122–3123 (1995)

[88] Ritchie H.: Application of the semi-Lagrangian method to a spectral model of the shallow water equations. Mon. Weather. Rev. **116**, 1587–1598 (1988)

[89] Ritchie H.: Implementation of the semi-Lagrangian method in a high resolution version of the ECMWF forecast model. Mon. Weather. Rev. **123**, 489–514 (1994)

[90] Ritchie H., Beaudoin C.: Approximations and Sensitivity Experiments with a Baroclinic Semi-Lagrangian Spectral Model. Mon. Weather. Rev. **122**, 2391–2399 (1994)

[91] Robert A.: The integration of a low order spectral form of the primitive meteorological equations. J. Meteorol. Soc. Jpn. **44**, 237–244 (1966)

[92] Robert A.: The integration of a spectral model of the atmosphere by the implicit method. Proc.of WMO/IUGG Symp. on NWP (1969)

[93] Robert A.: A stable numerical integration scheme for the primitive meteorological equations. Atmos. Ocean. **19**, 35–46 (1981)

[94] Rumley S., Nikolova D., Hendry R., Li Q., Calhoun D., Bergman K.: Silicon photonics for exascale systems. J. Lightwave Techn. **33**, 547–562 (2015)

[95] Sawyer J.S.: A semi-Lagrangian method of solving the vorticity advection equation. Tellus. **4**, 336–342 (1963)

[96] Simmons A., Hoskins B., Burridge D.: Stability of the semi-implicit method of time integration. Mon. Weather. Rev. **106**, 405–412 (1978)

[97] Simmons A., Burridge D.: An energy and angular-momentum conserving vertical finite-difference scheme and hybrid vertical coordinates. Mon. Weather. Rev. **109**, 758–766 (1981)

[98] Simmons A.J., Hollingsworth A.: Some aspects of the improvement in skill of numerical weather prediction. Q. J. R. Meteorol. Soc. **128**, 647–677 (2002)

[99] Skamarock W., Smolarkiewicz P., Klemp J.: Preconditioned Conjugate-Residual Solvers for Helmholtz Equations in Nonhydrostatic Models. Mon. Weather. Rev. **125**, 587–599 (1997)

[100] Skamarock W.: Evaluating mesoscale NWP models using kinetic energy spectra. Mon. Weather. Rev. **132**, 3019–3032 (2004)

[101] Skamarock W., Klemp J., Dudhia J., Gill D., Barker D., Duda M., Huang X., Wang W., Powers J.: A description of the Advanced Research WRF Version 3.pp.125 (2008)

[102] Smolarkiewicz P., Kühnlein C., Wedi N.: A consistent framework for discrete integrations of soundproof and compressible PDEs of atmospheric dynamics. J. Comput. Phys. **263**, 185–205 (2014)

[103] Smolarkiewicz P., Deconinck W., Hamrud M., Kühnlein C., Mozdzynski G., Szmelter J., Wedi N.: A finite-volume module for simulating global all-scale atmospheric flows. J. Comput. Phys. accepted for publication (2016)

[104] Staniforth A., Mitchell H.: A semi-implicit finite-element barotropic model. Mon. Weather. Rev. **105**, 154–169 (1977)

[105] Staniforth A., Temperton C.: Semi-Implicit Semi-Lagrangian Integration Schemes for a Barotropic Finite-Element Regional Model. Mon. Weather. Rev. **114**, 2078–2090 (1986)

[106] Staniforth A.: André Robert (1929-1993): His pioneering contributions to numerical modelling. Numerical methods in atmospheric and ocean modelling: The André J. Robert Memorial Volume. 25–54 (1995)

[107] Staniforth A., White A., Wood N., Thuburn J., Zerroukat M., Cordero E., Davies T., Diamantakis M.: Joy of UM 6.3 - Model Formulation. Unified Model documentation paper No 15 (2006)

[108] Staniforth A., Wood N.: Aspects of the dynamical core of a nonhydrostatic, deep-atmosphere, unified weather and climate-prediction model. J. Comput. Phys. **227**, 3445–3464 (2008)

[109] Staniforth A., Thuburn J.: Horizontal grids for global weather and climate prediction models: a review. Q. J. R. Meteorol. Soc. **138**, 1–26 (2012)

[110] Steppeler J., Hess R., Schättler U., Bonaventura L.: Review of numerical methods for nonhydrostatic weather prediction models. Meteorol. Atmos. Phys. **82**, 287–301 (2003)

[111] Staniforth A., Melvin T., Cotter C.: Analysis of a mixed finite element pair proposed for an atmospheric dynamical core. Q. J. R. Meteorol. Soc. **139**, 1239–1254 (2012)

[112] Taylor G.: The oscillations of the atmosphere. Proc. Roy. Soc. London. **A156**, 318–326 (1936)

[113] Temperton C., Staniforth A.: An efficient two-time level semi-Lagrangian semi-implicit integration scheme. Q. J. R. Meteorol. Soc. **113**, 1025–1039 (1987)

[114] Temperton C.: On scalar and vector transform methods for global spectral models. Mon. Weather. Rev. **119**, 1303–1307 (1991)

[115] Temperton C.: Treatment of the Coriolis terms in semi Lagrangian spectral models. ECMWF Technical Memorandum 213 (1995)

[116] Temperton C., Hortal M., Simmons A.: A two-time-level semi-Lagrangian global spectral model. Q. J. R. Meteorol. Soc. **127**, 111–127 (2001)

[117] Termonia P., Hamdi R.: Stability and accuracy of the physics-dynamics coupling in spectral models. Q. J. R. Meteorol. Soc. **133**, 1589–1604 (2007)

[118] Termonia P.: Study of the lateral boundary condition temporal resolution problem and a proposed solution by means of boundary error restarts. **137**, 3551–3566 (2009)

[119] Termonia P., Voitus F., Degrauwe D., Caluwaerts S., Hamdi R.: Application of Boyd's periodization and relaxation method in a spectral atmospheric limited-area model. Part 1: Implementation and Reproducibility tests. Mon. Weather Rev. **2012**, 3137–3148 (2012)

[120] Thuburn J., Staniforth A.: Conservation and linear Rossby mode dispersion on the spherical C-grid. Mon. Weather. Rev. **132**, 641–653 (2004)

[121] Thuburn J.: Some conservation issues for the dynamical cores of NWP and climate models. J. Comput. Phys. **227**, 3715–3730 (2008)

[122] Thuburn J.: Rossby wave dispersion on the C-grid. Atmos. Science Lett. **8**, 37–42 (2007)

[123] Tolstykh M.: Vorticity-divergence semi-Lagrangian shallow water model of the sphere based on compact finite differences. J. Comput. Phys. **179**, 180–200 (2002)

[124] Tolstykh M., Shashkin V.: Vorticity-divergence mass-conserving semi-Lagrangian shallow-water model using the reduced grid on the sphere. J. Comput. Phys. **231**, 4205–4233 (2012)

[125] Tolstykh M., Geleyn J.F., Volodin E., Bogoslovskii N., Vilfand R., Kiktev D., Krasjuk T., Kostrykin S., Mizyak V., Fadeev R., Shashkin V., Shlyaeva A., Ezau I., Yurova A.: Development of the multiscale version of the SL-AV global atmospheric model. Russ. Met. and Hydr. **40**, 374–382 (2015)

[126] Vaughan G., Worthington R.M.: Inertia-gravity waves observed by the UK MST radar. Q. J. R. Meteorol. Soc. **133**, 179–188 (2007)

[127] Venkatesh T., Mathew J.: The problem of clear air turbulence: changing perspectives in the understanding of the phenomenon. Sadhana-academy proceedings in engineering sciences **38**, 707–722 (2013)

[128] Wang H., Halpern P., Douglas J., Dupont T.: Numerical solutions of the one-dimensional primitive equations using Galerkin approximations with localized basis functions. Mon. Weather. Rev. **100**, 738–746 (1972)

[129] Warner T., Peterson R., Treadon R.: A tutorial on lateral boundary conditions as a basic and potentially serious limitation to regional numerical weather prediction. Bull. Amer. Meteor. Soc. **78**, 2599–2617 (1997)

[130] Wedi N.: Increasing horizontal resolution in numerical weather prediction and climate simulations: illusion or panacea. Phil. Trans. R. Soc. **372**, 1–12 (2014)

[131] Wedi N., Bauer P., Deconinck W., Diamantakis M., Hamrud M., Kühnlein C., Malardel S., Mogensen K., Mozdzynski G., Smolarkiewicz P.: The modelling infrastructure of the Integrated Forecasting System: Recent advances and future challenges. Technical Memorandum ECMWF 760 (2015)

[132] White A.A., Hoskins B.J., Roulstone I., Staniforth A.: Consistent approximate models of the global atmosphere: shallow, deep, hydrostatic, quasi-hydrostatic and non-hydrostatic. Q. J. R. Meteorol. Soc. **131**, 2081–2107 (2005)

[133] Wiin-Nielsen A., Brown J., Drake M.: On Atmospheric Energy Conversions Between the Zonal Flow and the Eddies. Tellus. **15**, 261–279 (1963)

[134] Williams R.T.: On the formulation of finite-element prediction models. Mon. Weather. Rev **109**, 463–466 (1981)

[135] Williamson D.: The evolution of dynamical cores for global atmospheric models. J. Meteorol. Soc. Jpn. **85B**, 241–269 (2007)

[136] Winninghoff F.J.: On the adjustment toward a geostrophic balance in a simple

primitive equation model with application to the problems of initialization and objective analysis. Phd thesis, University of California (1968)

[137] Wood N., Staniforth A., White A., Allen T., Diamantakis M., Gross M., Melvin T., Smith C., Vosper S., Zerroukat M., Thuburn J.: An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. Q. J. R. Meteorol. Soc. **140**, 1505–1520 (2014)

[138] Wyngaard J.: Toward numerical modeling in the "Terra Incognita". J. Atmos. Sci. **61**, 1816–1826 (2004)

[139] Yessad K., Wedi N.: The hydrostatic and non-hydrostatic global model IFS/ARPEGE: deep layer formulation and testing. Technical Memorandum ECMWF (2011)

[140] Zangl G., Reinert D., Ripodas P., Baldauf M. The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. Q. J. R. Meteorol. Soc. **141**, 563–579 (2015)