



UNIVERSITY OF BOLOGNA  
DEPARTMENT OF PHYSICS AND ASTRONOMY  
MASTER DEGREE IN SCIENCE OF CLIMATE

~ · ~

ACADEMIC YEAR 2024–2025

# CLIMATE SYSTEM MODELING

## NOTES

**Prof.**  
Antonio NAVARRA

**Student**  
Students of SOC  
00000

LAST UPDATED: December 13, 2024, [download latest](#)

The contents of these notes stem from the course of Climate System Modeling taught by Prof. **Antonio Navarra**<sup>1</sup> at University of Bologna for the academic year 2024/2025 and from his notes available at [wanderer.cmcc.it](http://wanderer.cmcc.it)<sup>2</sup>.

The source code is available on [Github](#).

---

<sup>1</sup>[CMCC, Unibo](#)

<sup>2</sup>If prompted, use the credentials: { user: cmcc , psw: pD8phJg3e76J }

# Contents

<b>1 Physical description of the climate system</b>	<b>7</b>
1.1 Atmosphere . . . . .	7
1.1.1 The General Circulation of the Atmosphere . . . . .	8
1.1.2 The time averaged zonal general circulation . . . . .	10
1.1.3 The horizontal general circulation and wind . . . . .	10
1.1.4 Mean Sea Level Pressure and The Sea Surface Temperature . . . . .	11
1.2 Oceans . . . . .	13
1.2.1 The salinity structure of the ocean . . . . .	14
1.2.2 Influence of the ocean . . . . .	17
1.3 Ice . . . . .	21
1.4 Interconnections between atmosphere, ocean and ice . . . . .	21
1.4.1 Atmosphere effect . . . . .	22
1.4.2 Ocean effect . . . . .	22
1.4.3 Ice . . . . .	23
<b>2 Fundamental Equations and processes</b>	<b>25</b>
2.1 Fundamental Equations . . . . .	25
2.1.1 Introduction . . . . .	25
2.1.2 Coordinate systems . . . . .	25
2.1.3 Equation of motion . . . . .	27
2.1.4 Summary of fundamental equations- . . . . .	31
2.1.5 Simplified equations- . . . . .	31
2.1.6 Linear solutions . . . . .	34
2.1.7 Waves . . . . .	34
2.2 Homogeneous flows . . . . .	35
2.2.1 The vorticity equation on the $\beta$ -plane . . . . .	38
2.3 Rossby waves . . . . .	39
2.4 Fundamental processes . . . . .	40
2.4.1 Radiation . . . . .	40
2.4.2 Moisture . . . . .	44
2.4.3 Clouds . . . . .	46
2.4.4 Surface Processes . . . . .	47
2.4.5 Hydrology . . . . .	48
<b>3 Numerical methods used in atmospheric models</b>	<b>49</b>
3.1 Historical introduction . . . . .	49
3.2 Methods for the numerical solution of the equations of motion . . . . .	50
3.3 Grid point method . . . . .	51
3.3.1 Finite difference schemes . . . . .	53
3.3.2 Convergence . . . . .	55
3.3.3 Stability . . . . .	56
3.4 Time differencing schemes . . . . .	60

3.4.1	Properties of schemes applied to particular equations . . . . .	63
3.4.2	A combination of schemes . . . . .	68
3.5	The Advection Equation . . . . .	68
3.6	Energy and numerical methods . . . . .	72
3.6.1	Conservation in numerical schemes . . . . .	74
3.6.2	The Arakawa Jacobian . . . . .	75
3.6.3	Gravity waves . . . . .	76
3.6.4	Two dimension . . . . .	79
3.7	Spectral Methods . . . . .	82
3.7.1	Numerical spectral solution (nonlinear) . . . . .	85
3.8	Spherical frame . . . . .	86
3.8.1	The Helmholtz theorem . . . . .	88
3.8.2	Spherical Harmonics . . . . .	89
3.8.3	Spectral Transforms . . . . .	90
3.8.4	Truncation . . . . .	93
3.8.5	Summary . . . . .	93
3.9	Lagrangian Methods . . . . .	97
3.9.1	Semi-Lagrangian Methods . . . . .	98
3.10	Vertical coordinates . . . . .	99
<b>4</b>	<b>Climate Models</b>	<b>101</b>
4.1	Radiative Model . . . . .	101
4.1.1	Two-layer atmosphere . . . . .	104
4.1.2	Three-layer atmosphere . . . . .	105
4.2	Radiative-Convective Model . . . . .	107
4.2.1	Radiative convective equilibrium . . . . .	107
4.3	Radiative forcing . . . . .	109
4.3.1	Greenhouse feedback . . . . .	111
4.3.2	Water vapor feedback . . . . .	112
<b>5</b>	<b>Ocean models</b>	<b>117</b>
5.1	Ice models . . . . .	120
5.2	The River Transport . . . . .	120
5.3	Grids . . . . .	120
5.3.1	Regular grids . . . . .	120
5.3.2	Resolution and mesoscale eddies . . . . .	123
5.4	Caos . . . . .	125
5.5	Vorticity dynamics . . . . .	126
5.5.1	What maintains the circulation? . . . . .	126
<b>6</b>	<b>Exploring Variability within Models</b>	<b>129</b>
6.1	Designing a model for variability . . . . .	132
6.2	General Circulation Models . . . . .	135
<b>7</b>	<b>Forcings</b>	<b>139</b>
7.1	SST forcing . . . . .	139
7.1.1	Climate forcings on SST . . . . .	139
7.1.2	How can surface sea temperature impact atmospheric circulation? . . . . .	140
7.1.3	What is the mechanism by which SST is affecting climate? . . . . .	144
7.2	Steady heating forcing balances . . . . .	145
7.2.1	Balancing advection . . . . .	145
7.2.2	Balancing vorticity . . . . .	147

7.3	Atmosphere behaviour: linear or non-linear? . . . . .	147
7.4	Potential predictability . . . . .	148
7.4.1	ANOVA . . . . .	149
7.4.2	Challenges in observations . . . . .	151
<b>8</b>	<b>Earth System Models</b>	<b>153</b>
8.1	Expectations with increasing greenhouse gases . . . . .	154
8.2	Expanding the complexity of climate modeling . . . . .	155
8.3	Carbon and Biogeochemical cycles . . . . .	156
8.3.1	The Carbon cycle . . . . .	157
8.3.2	Complexity and model evaluation . . . . .	159
8.4	Digital Twins . . . . .	162
8.4.1	Enabling Technologies . . . . .	162
8.5	Climate Ensembles . . . . .	163
<b>9</b>	<b>Theory of symmetric circulation</b>	<b>165</b>
<b>10</b>	<b>Tipping points</b>	<b>171</b>
10.1	Examples of Climate System Tipping Points . . . . .	172
10.2	Climate impacts . . . . .	174
<b>11</b>	<b>Are General Circulation models obsolete?</b>	<b>177</b>
<b>A</b>	<b>Appendix</b>	<b>179</b>
A.1	Spherical harmonics expansion . . . . .	179
A.2	Vector calculus . . . . .	182
A.2.1	Curl . . . . .	182
A.2.2	Gradient . . . . .	182
A.2.3	Divergence . . . . .	182
A.3	EOF: Empirical Orthogonal Functions . . . . .	183
A.3.1	Eigenvalue decomposition → Singular Value Decomposition . . . . .	184
A.3.2	Limitations . . . . .	185
A.3.3	Variance . . . . .	185
A.3.4	Covariance . . . . .	185
A.3.5	Correlation . . . . .	185
A.4	Space-time splitting . . . . .	185
A.4.1	Zonal mean . . . . .	185
A.4.2	Time mean . . . . .	186
A.4.3	Higher order quantities . . . . .	186
	<b>Bibliography</b>	<b>189</b>



# Chapter 1

## Physical description of the climate system

Last updated: 2024-12-13

Source file: [chapter-intro.tex](#)

### 1.1 Atmosphere

#### Composition

The atmosphere is a gaseous envelope that surrounds the Earth, acting as a protective envelope essential for life. It is composed of a mechanical mixture of several gases, with nitrogen being the most abundant (78% of the volume of dry air), followed by oxygen (20.95%), argon (0.93%), and carbon dioxide (0.037%). Trace gases like neon, helium, methane, and hydrogen also contribute to its composition, along with varying amounts of water vapor, which depend on factors such as location, evaporation rates, and temperature.

This gaseous mixture is held to the Earth by gravity, which causes the atmosphere to exert pressure. Near the surface, the pressure is greatest, approximately  $1.013 \times 10^3$  mb at sea level, and it decreases with altitude. This reduction in pressure reflects the compressibility of gases, making the atmosphere densest near the ground. This characteristic distinguishes the atmosphere from the ocean, where the density remains relatively uniform due to the near incompressibility of liquids.

#### The Vertical Temperature Gradient

The atmosphere exhibits a distinct temperature profile that changes with altitude and varies depending on the region and season. These variations define the layers of the atmosphere, each with unique characteristics and processes.

- **The Troposphere:** is the lowest layer, extending from the surface up to about 17 kilometers in the tropics and 8–9 kilometers in polar regions. In this layer, the temperature generally decreases with altitude at an average rate of  $6.4^\circ\text{C km}^{-1}$ . This gradient arises because the surface absorbs heat from the sun and transfers it to the air above, while higher altitudes are farther from this heat source. The troposphere is also the region where most weather phenomena, including clouds and storms, occur, making it the most dynamic part of the atmosphere.

At the upper boundary of the troposphere, known as the **tropopause**, the temperature reaches its minimum and remains constant for a short distance. In tropical regions, the tropopause is particularly cold, with temperatures dropping to about 190 K, marking one of the coldest points in the Earth's atmosphere.

- **The Stratosphere:** Above the tropopause lies the stratosphere, a more stable layer where temperatures either remain constant or begin to rise with altitude. This increase is primarily due to the absorption of ultraviolet radiation by ozone molecules, which warms this part of the atmosphere. Unlike the troposphere, the stratosphere lacks the turbulence and weather systems associated with surface heating.
- **The Thermosphere:** the temperature continues to increase steadily with altitude. This layer is less influenced by surface-level processes and is instead heated by the absorption of high-energy solar radiation, making it the hottest layer of the atmosphere.

	Troposphere (sum/win)	Tropopause	Stratosphere
poles	280 / 235	9 km: 230 / inversion (1.5/2 km)	Slow Warming/ slow cooling
midlat	290 / 280	10 km: 220	Slow Warming/ slow cooling
tropics	300	17 km: 190	warming

## Regional and Seasonal Variations

Temperature profiles in the atmosphere are not uniform and vary significantly between regions and seasons.

- **Polar regions:** In summer, temperatures near the surface reach about 280 K, decreasing gradually with height. In winter, surface temperatures can plummet to 235 K, but a slight increase often occurs at altitudes of 1.5–2 kilometers due to a phenomenon called temperature inversion. This inversion, caused by the cooling of surface air, reverses the usual trend of decreasing temperature with height.
- **Tropical regions:** Surface temperatures remain relatively constant, averaging around 300 K. However, as altitude increases, the temperature drops sharply, reaching the coldest point at the tropopause (about 190 K).
- **Mid-latitudes:** The temperature profile combines characteristics of both polar and tropical regions, varying more noticeably between seasons.

### 1.1.1 The General Circulation of the Atmosphere

Since heated air tends to rise and low-level air must flow in to replace the risen heated air, the relative heating and cooling of different areas in the Earth's surface plays a significant role in driving local winds and the large-scale atmospheric circulation. The atmospheric circulation can be summed up into two different contributes: the east-west motion of the *Walker Circulation* and the north-south circulation of the three-cell structure: Hadley, Ferrel and polar cells. The Coriolis effect plays a crucial role in wind deflection to the right in the Northern Hemisphere and to the left in the Southern Hemisphere explaining the formation of trade winds, westerlies, and polar easterlies.

1. - **The Walker circulation** refers to the equatorial Pacific and involves rising air over the region of Indonesia and descending air over the eastern Pacific. It was examined following the discovery of strong negative correlation between surface pressure anomalies in the two regions. Variations in the strength of this circulation produce a large scale fluctuation with an irregular period, known as Southern Oscillation.

#### 2. Three-cell pattern

*The tropical Hadley cell* is driven by solar heating, causing rising motion near the Equator, then by the release of latent heat as the rising air leads to precipitation in the

rising-air branch of the cell. This rising-air branch of the Hadley cell is not centered consistently on the equator but migrates north & south. Moreover, the strength of the Hadley circulation varies with longitude, being strongly affected by such factors as whether the underlying surface is land or ocean. After rising near the Equator, the air in the Hadley cell moves poleward, sinking near  $30^{\circ}\text{N}$  and  $30^{\circ}\text{S}$  and thereby generating belts of surface-level high pressure near  $30^{\circ}\text{N}$  and  $30^{\circ}\text{S}$ . Since the sea level pressure is low, the high pressure produced by the sinking air at about  $30^{\circ}\text{N}$  and  $30^{\circ}\text{S}$  creates a surface pressure gradient leading to the movement of a portion of the sinking air back toward the low pressure near the equator. The region of low-level convergence toward the bottom of the rising-air branch of the Hadley cell is called the intertropical convergence zone (ITCZ). The position of the ITCZ shifts with the seasons, moving north during the Northern Hemisphere's summer and south during its winter. This seasonal movement affects global wind and rainfall patterns, contributing to phenomena like monsoons.

*In the mid-latitude Ferrel cell*, termed indirectly because of having rising air in its cooler branch, the low-level flow is toward the poles, away from the relatively high pressure produced by the descending arms of the Hadley and Ferrel cells at about  $30^{\circ}$  latitude and toward the relatively low pressure at about  $60^{\circ}$  latitude.

*The polar cell*: strong radiational cooling near the poles, causes polar air to become cold and dense, which in turn causes it to sink. Thus there is relatively high pressure at the pole, which, combined with the low pressure near  $60^{\circ}\text{N}$  and  $60^{\circ}\text{S}$  discussed in connection with the rising-air branch of the Ferrel cell, produces surface flow equatorward from the pole. This polar cell is extremely weak, although it remains detectable in time averages of the air circulation.

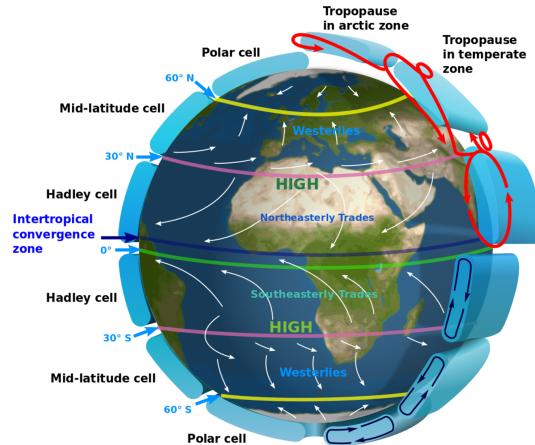


Figure 1.1: Global Atmospheric Circulation cells

In reality, the atmospheric circulation is more complex than the three-cell model suggests. It exhibits significant temporal and spatial variations due to factors like land/sea contrasts, topography, and changes in solar heating. While the model serves as a foundational framework, modern climate science relies on satellite data and simulations for a detailed understanding of global atmospheric dynamics.

**Monsoons and Land Sea breezes:** they are wind patterns driven by differences in heating between land and water surfaces, but they operate on different scales.

**Sea Breeze (Daytime):** During the day, land heats up faster than water, causing air over the land to rise and creating a low-pressure zone. Cooler, denser air from over the sea flows in to replace it, generating a breeze from sea to land.

**Land Breeze (Nighttime):** At night, the land cools faster than water, creating a high-pressure

zone over the land. Air flows from the land to the sea, forming a weaker land breeze.

**Summer Monsoon:** Similar to a large-scale sea breeze, during summer, land masses (like South Asia) heat up more quickly than the surrounding oceans. This creates a low-pressure zone over land, drawing in moist air from the Indian or Pacific Oceans. The moisture-laden air rises and cools, resulting in heavy rainfall, particularly over mountain ranges like the Himalayas and Ghats.

**Winter Monsoon:** In winter, the land cools faster than the ocean, forming a high-pressure zone. Cold, dry air flows outward from the land to the sea, leading to dry weather conditions over most of South Asia.

**Key Differences:** While land/sea breezes are localized and operate on a daily cycle, monsoons occur over larger areas and are seasonal, driven by shifts in the Intertropical Convergence Zone (ITCZ). Monsoons involve additional complexities, including the influence of upper atmospheric circulations, such as the jet stream. The Asian monsoon, particularly over South Asia, is the most prominent example, crucial for regional agriculture and ecosystems. However, its dynamics are far more complex than those of simple land and sea breezes.

### 1.1.2 The time averaged zonal general circulation

The time-averaged circulation, computed from ERA5 Reanalysis data, shows westerly mid-atmosphere jets in the subtropical regions of both hemispheres. These jets reach a maximum around the 200mb level (12 km) and extend to the surface in the mid-latitudes, while easterly flows dominate the equatorial zone. The jets exhibit a strong seasonal cycle, with accelerated jets in winter and weaker jets in summer. There is a seasonal poleward migration of the jet cores, with the winter season showing more concentrated and intense maximums. The Southern Hemisphere's winter jet is broader than its Northern Hemisphere counterpart, both linked to the stratospheric flow above.

The stratosphere also exhibits strong jets with a stronger seasonal cycle, where easterlies replace westerlies as the seasons change. In the meridional circulation, low-level convergence occurs at the equator, with high-level divergence. The InterTropical Convergence Zone (ITCZ) oscillates between 15°N and 5°S, influenced by the seasonal cycle of the sun.

Temperature decreases with latitude due to radiation balance, with the equator receiving more solar radiation than the poles. The temperature decreases with height in the atmosphere, though the lapse rate is less than adiabatic, indicating a generally stable atmosphere. There is a strong latitudinal temperature gradient, which weakens with altitude and reverses in the upper atmosphere and stratosphere, due to radiation absorption by ozone and other components in the lower stratosphere.

The reversal of the meridional temperature gradient aligns with the zonal wind structure, showing positive shear in the troposphere and negative shear above, consistent with thermal wind balance. Water vapor, measured as specific humidity, is concentrated in the lower atmosphere and the equatorial region, with significant dryness at higher altitudes. The atmosphere's moisture content decreases with altitude due to the Clausius-Clapeyron relation, which links water vapor pressure to temperature. At the equator, moist air contains 12-15 grams of water per kilogram of air.

Specific humidity follows the seasonal cycle of the sun, shifting latitudinally. The maximum is located around 5°S in December-February (DJF) and around 10°N in June-July-August (JJA). The winter hemisphere is more moist at the surface than the summer, though values in the winter are smaller than in the equatorial zone, reaching up to 8 g/kg.

### 1.1.3 The horizontal general circulation and wind

The climatological geopotential height at 200mb shows deviations from a zonally symmetric circulation, particularly over the east coasts of continents, downstream from major mountain

ranges like the Rockies and Himalayas. These features are clearer with specialized projections, highlighting the winter atmospheric pattern. While geopotential height can approximate wind flow in mid-latitudes, wind analysis is more useful in low latitudes.

At 200mb, winter jet streams are strong in both hemispheres, especially over Asia, with the Asian jet reaching velocities over 70 m/s. The Southern Hemisphere jet is weaker. In summer, westerly jets are weaker, and easterly jets dominate the tropics. The Indian Ocean sees a strong easterly flow in summer, leading to divergence over South America and Indonesia.

The meridional wind shows alternating poleward and equatorward patterns in winter, with large deviations in the Pacific-North American sector and East Asia. In summer, winds weaken but remain stronger in the Southern Hemisphere. Near the surface, the zonal wind is stronger over oceans and weaker over land, with easterlies in the subtropics, except in the Indonesian region, where westerlies prevail. The equatorial Pacific shows a convergence zone.

During JJA, the seasonal cycle strengthens winter wind features, with strong westerlies in the Indian Ocean signaling the start of the South Asian Monsoon. The meridional wind shows an equatorward flow along continent coasts, intensifying in summer, and reversing along the Somali coast.

At the 850mb level, trade winds dominate the equatorial Pacific, shifting with the ITCZ's seasonal cycle. The Asian Summer Monsoon is visible in the Indian Ocean, forming a large gyre from East Africa to the Indian subcontinent, crucial for understanding low-level atmospheric and oceanic interactions.

#### 1.1.4 Mean Sea Level Pressure and The Sea Surface Temperature

(MSLP) is a key parameter in describing atmospheric circulation. It represents the atmospheric pressure adjusted to mean sea level, though its usefulness is limited in areas with significant mountains, like the Rockies, Himalayas, and Antarctica, where the concept of "sea level" is effectively underground. Outside these areas, MSLP provides a good representation of atmospheric mass distribution. High MSLP areas correspond to mass accumulation, particularly in the subtropics, where high-pressure systems are common in both summer and winter.

Near-surface temperatures, typically measured 2 meters above the ground, reflect the seasonal and geographical variation in temperature. Over oceans, they follow sea surface temperature (SST), but land areas show significant seasonal changes. Northern continents are colder than the oceans at the same latitude, and coastal regions also experience temperature differences, with west coasts being milder than east coasts. In winter, high pressure tends to remain over land while low pressure develops over the oceans, reversing in summer. The intertropical zone sees high-pressure centers shifting with the seasons, while the Southern Hemisphere features a more symmetric pattern, with a ring of low pressure around Antarctica.

The seasonal cycle is evident in the shift of high-pressure areas in the tropics, moving latitudinally with the changing seasons. The meridional distribution of sea level pressure shows high pressure in the tropics, with low pressure areas near the equator and in the mid-latitudes. In the Southern Hemisphere, the pressure gradients are stronger and more pronounced, with tropical pressure maxima shifting seasonally.

Sea Surface Temperature (SST) shows strong north-south gradients, with polar regions being cold and the equator generally warm. However, notable deviations from zonal symmetry occur near continental east coasts, particularly in the equatorial Pacific, where cold water intrudes at the equator in the East Pacific, contrasting with the warm waters of the West Pacific. These temperature differences highlight significant gradients in SST along the equator

## The Role of Energy Processes

The vertical temperature gradient in the atmosphere is maintained by complex energy processes, including the absorption and radiation of heat, convection, and interactions with the Earth's surface. These processes vary with latitude and season, creating the dynamic and layered structure of the atmosphere.

Inversions, which occur when surface air cools significantly, disrupt the usual temperature gradient, particularly in the troposphere. These events highlight the intricate balance of energy transfer within the atmosphere and its impact on weather and climate.

### Energy balances

To understand how the atmosphere maintains the various temperature structures, it is necessary to consider the processes by which the atmosphere gains and loses energy and how these processes vary with latitude and season. Essentially all the energy that enters in the climate system comes from the sun, part of it is absorbed in the system and must be balanced with outgoing energy leaving the system, to maintain the overall climate system in its observed equilibrium state.

Because of the different temperatures between the sun and the Earth, the energy emitted by the two bodies is sharply different. The spatial and temporal distribution of the receipt of solar energy at the Earth's surface is highly dependent upon the Earth's annual revolution around the sun, its daily rotation about its axis, and the tilt of its axis concerning the plane of its orbit.

When the Earth is in the portion of its orbit where the Northern Hemisphere is tilted toward the sun, the Northern Hemisphere receives the majority of the direct sunlight and experiences summer, while the Southern Hemisphere experiences winter. Naturally, the reverse occurs when the Earth is in the opposite portion of its orbit, with the Southern Hemisphere tilted toward the sun.

About half of the incoming solar radiation is absorbed by the Earth's surface. This energy is transferred to the atmosphere by warming the air in contact with the surface (sensible heat), by evaporation and by thermal radiation that is absorbed by clouds and by greenhouse gases. The atmosphere in turn radiates thermal radiation back to Earth as well as out to space. Since all the incoming radiation equals to  $342 \text{ Wm}^{-2}$  and  $107 \text{ Wm}^{-2}$  of it is reflected back to space, there must be  $235 \text{ Wm}^{-2}$  of terrestrial radiation emitted to space for an equilibrium situation from three main sources: water vapor,  $\text{CO}_2$ , clouds and Earth's surface.

At the top of the atmosphere, the energy balance is composed of the compensating fluxes between the incoming solar radiation and the outgoing thermal radiation from the Earth. The solar radiation will be modulated by the reflectivity caused by the atmospheric cloud and in smaller part by the molecular components of the atmosphere itself. The thermal radiation will be modulated by the absorption properties of the atmosphere and its constituents. Both will be affected by the circulation and ultimately by the atmospheric flows.

The net balance shows a surplus of radiative flux in the subtropical region and a deficit in the polar regions. The global radiation budget can be obtained by integrating over the surface of the Earth the radiative fluxes.

At the surface, the energy balance involves more processes. There is also here a net solar radiation flux, modulated by the albedo of the surface and of the atmosphere, and a net thermal flux, obtained as the balance between the radiation emitted by the surface and the downward flux coming from the bulk of the atmosphere. Then there is a sensible heat flux caused by the turbulent vertical motion that carries away heat through mechanical agitation and the latent heat flux that represents the heat necessary for the evaporation of water from the surface. Because of the extent of the ocean, the latent heat flux is an important element of the budget.

In addition, we can look at the total cloud cover which is essentially the aggregated quantity of clouds at various levels. This is the quantity that is intuitively linked to the observation of a “cloudy sky”. In general, clouds are present over a vast surface of the globe. Some areas show a persistent absence of cloudiness over the entire year, such as the Sahara desert, part of the Arabian peninsula, Australia, and South Africa. Other areas instead show a strong seasonal cycle, with different cloud cover in different seasons. It is interesting to note that the seasonal cycle takes different characteristics in different areas. In the mid-latitudes cloud cover is larger in the local hemispheric winter than in the summer.

Looking at the zonal profile of the clouds, it is evident that major concentrations are related to the equator and middle/high latitudes.

## 1.2 Oceans

Oceans, covering approximately 71% of Earth’s surface, play a crucial role in global climate systems by storing and transferring heat, nutrients, and momentum. The interaction between the atmosphere and oceans significantly impacts weather and climate patterns. These bodies of water contain an estimated 1,350 million cubic kilometers of water, predominantly found in the Pacific, Atlantic, Arctic, and Southern Oceans, with an average depth of about 4,000 meters.

Surrounding continents are continental shelves, shallow areas that extend outward for several kilometers before dropping sharply at the continental slope into the deep ocean floor. The ocean floor itself is varied, featuring features like underwater mountains, ridges, trenches, and smooth sedimentary basins.

The oceans redistribute solar energy from equatorial to higher latitudes. As water warms at low latitudes, it absorbs significant amounts of solar heat. This heat is released into the atmosphere in the form of latent heat and longwave radiation as water cools or condenses, influencing global atmospheric circulation. Ocean currents further distribute this heat, balancing regional temperatures and aiding in climate regulation.(\*)

**Composition.** Seawater is not just water but a mixture containing various dissolved salts, with an average salinity of 35 grams per kilogram of water. This salinity largely comes from chloride, sodium, and sulfate, accounting for the majority of dissolved material. Salinity levels, however, vary globally. The Arctic Ocean shows lower salinity ( 29%), influenced by melting ice and freshwater inflows, while the subtropical Atlantic has some of the highest salinities ( 37.5%), due to high evaporation and low precipitation. These variations affect density, which in turn drives ocean circulation and temperature dynamics. Overall, oceans are not only a storage medium for heat and nutrients but also critical in redistributing energy across the planet, moderating climate, and supporting ecosystems. Their interaction with atmospheric processes underscores their integral role in Earth’s environmental systems.

### Temperatures profiles.

**Surface Temperatures:** Range: -1°C (polar regions) to 20–30°C (tropics). Seasonality: temperatures align in east-west zonal patterns, with exceptions like:

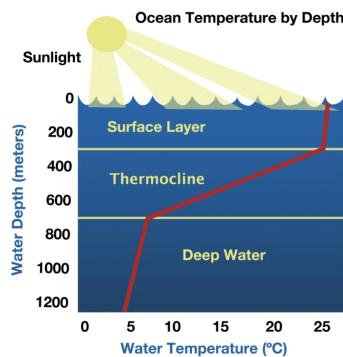
- Tropical Pacific: western regions warmer than eastern regions due to currents.
- Gulf Stream: transports warm water northeast, moderating European climates.

### Temperature with Depth:

- Mixed Layer (0~30 m in summer): warmed by sunlight and mixed by winds.

- Thermocline ( $200\text{--}1000$  m): sharp temperature gradient separating surface water from deeper layers.
- Deep Ocean ( $> 1000$  m): uniform cold temperatures ( $0.5\text{--}1.25^\circ\text{C}$ ).
- Coldest water ( $-0.25^\circ\text{C}$ ) near Antarctica due to sinking dense, salty water from surface cooling and sea ice formation.

R0.5



Ocean circulation is driven by temperature and salinity differences, influencing density and pressure. Surface currents are wind-driven, forming large gyres such as those in the Pacific and Atlantic Oceans. These gyres redistribute heat and nutrients, shaping regional climates and ecosystems. Deep ocean currents, part of the thermohaline circulation<sup>1</sup>, involve the sinking of dense, cold water and the upwelling of warmer, nutrient-rich water, critical for sustaining marine life. Additionally, mesoscale eddies, swirling water masses, play a significant role in transporting momentum, heat, and nutrients horizontally and vertically, although their dynamics are not fully understood. These processes together illustrate the complex and vital role oceans play in regulating Earth's climate and supporting life.

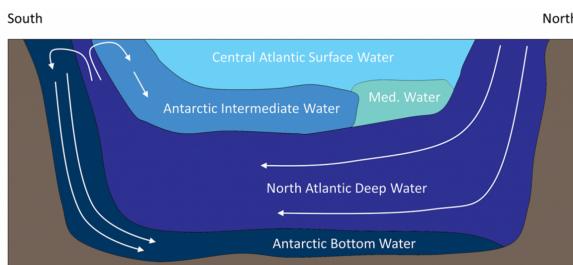


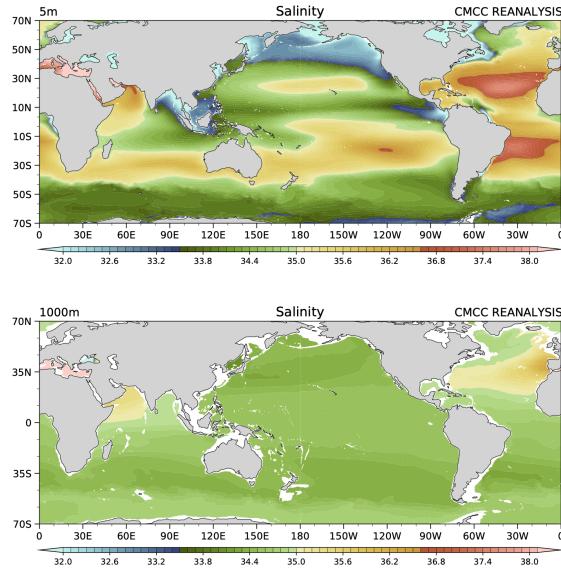
Figure 1.2: The major water masses of the Atlantic Ocean

### 1.2.1 The salinity structure of the ocean

The salinity structure of the oceans is shown in Figure 1.2.1. The top panel shows the salinity near the surface at a nominal depth of 5m. We notice that there is a complex structure with low salinity (fresher) waters at the poles and progressively more saline water moving towards the Equator, but then salinity decreases again at the equator. It is probably instructive to look at the latitudinal distribution of the precipitation. We can notice that the peaks of precipitation in the midlatitude and at the Equator are correlated with the low-salinity areas, whereas the subtropical regions are regions of strong evaporation, whose signature is the high salinity of the surface waters.

<sup>1</sup><https://rwu.pressbooks.pub/webboceanography/chapter/9-8-thermohaline-circulation/>

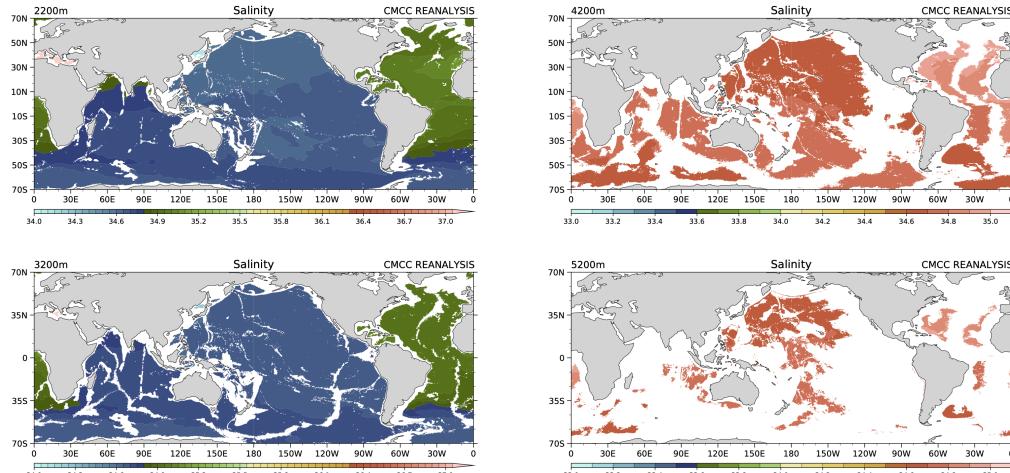
The Pacific Ocean is fresher than the Atlantic or even the Indian Ocean. We can see from the map at 1000m that sources of saline waters are marginal seas like the Mediterranean and the Persian Gulf. These areas are evaporative basins that produce water so saline that it dominates over the temperature effect and becomes denser so that we can find Mediterranean water below the surface. R0.5



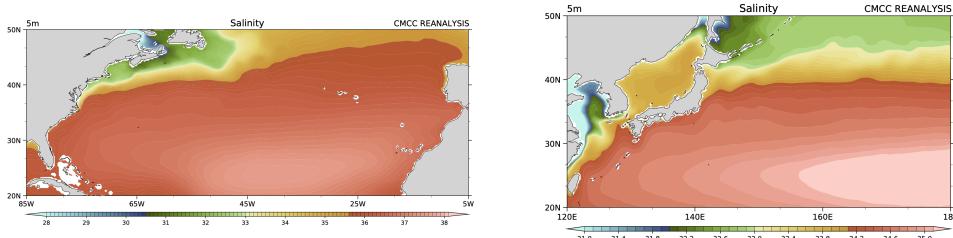
The effect of the precipitation is less visible at the deeper depth pf 1000m (bottom panel), where the ocean tends to be fresher and more uniform. We are using here the same scale to give a feeling of the changes in salinity, except for the Atlantic and the east Indian Ocean, the salinity is around 34 psu. The effect of the runoff from major river systems is visible along the Atlantic coast of South America where the Amazon river discharges and in the Bay of Bengal and around Indochina from the run-off the major rivers there, the Ganges and the Mekong.

In the pictures below: on the left salinity deviations with depth, going even deeper we have to change drastically the scale. The oceans are remarkably uniform and the salinity deviations are really small. On the right, Pacific and Atlantic ocean salinity.

Looking at the North Atlantic (figures below) we notice that there is a strong salinity gradient along the North American Coast that follows roughly the pattern of the temperature gradient shown below. Strong temperature gradients are presumably to be connected to the existence of currents, but we will need to check the density, depending on the salinity later, to be sure.



Anyway, this picture is giving a strong indication of the existence of something remarkable and intense along the western boundary of the Atlantic ocean.



A similar situation exist in the North Pacific along the Japan coast and therefore we can start to suspect that this has to do with the presence of the continental boundary.

The previous analysis of the temperature is giving us hints of a strong vertical structure of the oceans, so it may be useful to look at the vertical distribution somewhat more in detail. The figures above show the same section North-South section of along the longitude of 25W, roughly in the middle of the Atlantic Ocean. The salinity follows a similar pattern as the temperature with a strong gradient approximately at the thermocline, the high salinity is confined in thin; layers at the surface, but some interesting behaviour is visible below. In the Northern Hemisphere we can see the Mediterranean water penetrates at depth and actually protruding under the fresher water of Antarctic origin that is colder, but because is fresher floats over the Mediterranean water. Really cold Antarctic water reaches the bottom, filling the abyssal plains of the basin.

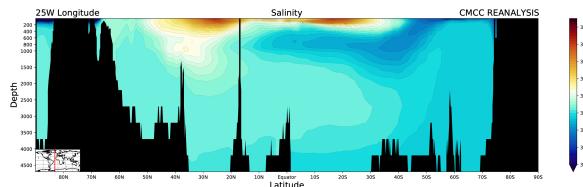


Figure 1.3: North-South Salinity section at 25W longitude.

The Mediterranean waters are clearly visible in a longitude-depth section (Figure 1.3). The saline mediterranean water sinks to about 1000m because it is warmer than the Atlantic but much more saline so it is denser and it reaches an equilibrium depth at about 1000m.

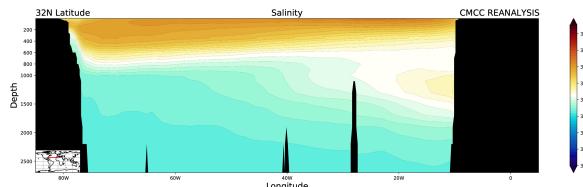


Figure 1.4: The vertical structure of the Pacific Ocean

The vertical structure of the Pacific Ocean (Figure 1.4) is different and what we see is a situation where salinity is slowly varying getting fresher toward the surface, with the thin saline water in the subtropics a clear signature of the equatorial upwelling and the Antarctic water filling the abyssal plains. The Indian Ocean is similar to the Pacific South of the Equator but North of the Equator at this longitude is fresher.

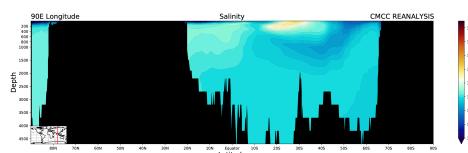


Figure 1.5: As in Fig. 1.4 but for 90E longitude

The Indian Ocean (Figure 1.5) shown here as a section cutting essentially through the Bay of Bengal, shows a similar strong stratification, but there are only weak signs of an equatorial upwelling of cold water.

We can gain more insights in the equatorial structure by looking at a longitudinal section along the Equator in the Pacific (Figure ??). Here we see the general difference between the Atlantic and the Pacific, but we also see that the salinity is following the slant of the equatorial thermocline and the equatorial upwelling in the East Pacific. The effect of the major precipitation center of ITCZ in the West Pacific is visible in fresh water at the surface in the West.

### 1.2.2 Influence of the ocean

**The continentality effect** It makes continental summers warmer and winters cooler than the adjacent ocean areas. The oceans have this effect because of their large heat storage capacity and because part of the energy that they retain from solar input during the summer they return to the atmosphere in the forms of sensible and latent heat and longwave radiation the following winter. Conversely, the continents have the reverse effect, strengthening the seasonal temperature contrasts

**Ocean currents** They can transport heat from low to high latitudes. (in the Gulf Stream of the North Atlantic). This contributes to much higher winter air temperatures over mid-to-high latitude oceans than over adjacent land areas at the same latitude.

**The overall basin circulation** The surface currents of the Atlantic ocean are shown in Figure 1.6. As we might have suspected from the temperature coast that reaches all the way across the North Atlantic to Europe, the Gulf Stream. Strong currents are also visible in the Equatorial area where they connect to the mid-latitude circulation forming a large circular system, the Subtropical Gyre. The South Atlantic has a similar gyre in the subtropical region, but at higher latitudes we can notice a strong westerly current cutting all along the basin, essentially along a latitude line between 40S and 50S.

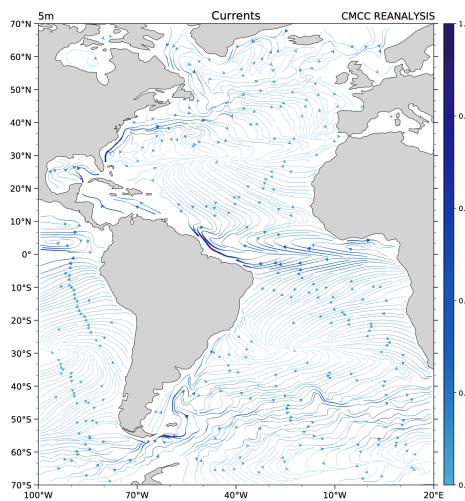


Figure 1.6: The surface currents of the Atlantic ocean

The North Pacific surface circulation is shown in Figure 1.7. We can notice here again a strong current along the Western boundary of the basin that than feed into a basin wide gyre that connects to the equatorial circulation. The boundary current, known her as the Kuroshio Current, is very narrow and intense along the Japan coast, as it is also the case of

the Gulf Stream in the Atlantic, and it tapers into a wide system of streams and eddies into the open ocean.

It is possible to see also local system, like the small gyre off the Alaskan coast and similar circulation in the marginal seas, like the Sea of Okhotsk, near the Siberian coast. Their presence is remarkable as we are looking here at climatological averages over more than 40 years and so they are stable and persistent features. This is another reminder of how even relatively smaller feature in the ocean can be climatologically persistent over many years.

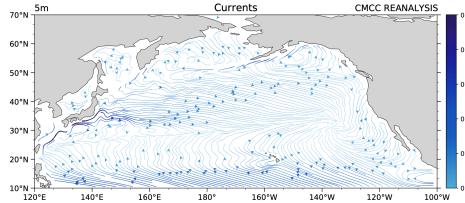


Figure 1.7: The North Pacific surface circulation

The circulation of the South Pacific Ocean is shown in Figure 1.8. The subtropical gyre is visible also here, but there is a weak indication of a western intensification current in the West pacific, close to the coasts of Australia and New Zealand. The strong high latitude current that we have seen in the Atlantic is also present here and evidently connect to the other basin through the Drake Passage, that is the Straits between South America and Antarctica.

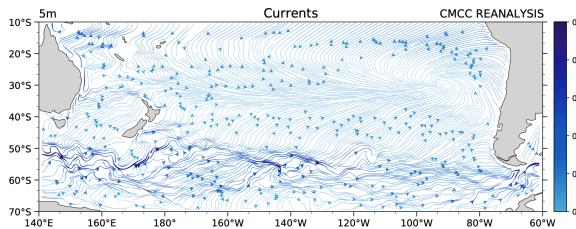


Figure 1.8: The circulation of the South Pacific Ocean

The Indian Ocean circulation is shown in Figure 1.9. The Indian Ocean is confined by continental masses in the north, so it is mostly composed of the equatorial region and the midlatitudes are all in the Southern Hemisphere. The subtropical gyre is present, together with features North of Madagascar. A boundary current develops on the African coast, The westerly current in the southern mid-latitudes is visible also here, strong and with a vigorous eddy field.

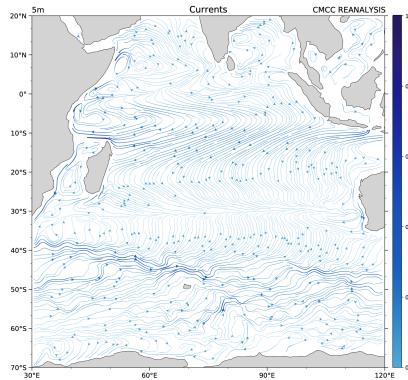


Figure 1.9: The Indian Ocean circulation

At this point we can suspect that there is probably a continuous ring of currents around

Antarctica and this can be confirmed by the bottom panel of Figure 1.9 that shows the entire extent around a longitude circle of the current, known as the Antarctic Circumpolar Current. It is a strong, highly turbulent system that connects all the Ocean Basins.

**The equatorial circulation** The Equator is a special place for the atmosphere and so it is a special place also for the Oceans. The circulation in this area is strongly coupled with the atmospheric circulation and it is often characterized by both westerly and easterly currents and by special behaviours right at the Equator line. Furthermore, it is also very different from ocean to ocean.

The equatorial current system in the Pacific Ocean is shown in Figure 1.10. It is a complex system composed of two easterly currents, the North and South Equatorial currents, sandwiching the westerly Equatorial countercurrent. It is possible to notice that at the Equator the currents are strongly easterly and diverging, leading to the emergence of upwelling at the Equator by Ekman transport.

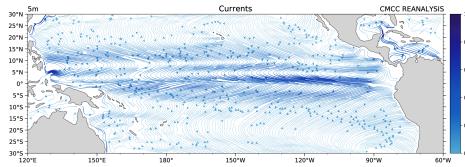


Figure 1.10: The equatorial current system in the Pacific Ocean

The complexity of the equatorial current system can be further appreciated by looking at the vertical section of the zonal current at the Equator. A strong westerly current below the surface, slanting from the West Pacific to the East Pacific, is visible at depth between 200 and 300m. The speed is in excess of 1 m/s and it gets progressively shallower in the East.

It is however part of an alternation of westerly and easterly currents that become progressively weaker as they get deeper. They are centered almost perfectly at the Equator, as it can be seen in Figure 1.11. The latitudinal position of the undercurrent is very tightly controlled by rotational effects and it precisely tracks the position of the equatorial line.

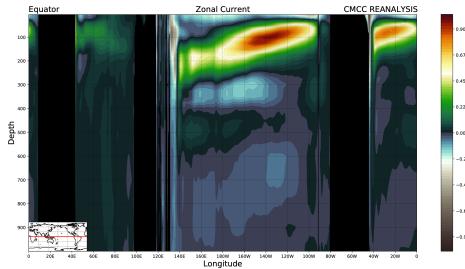


Figure 1.11: Alternation of westerly and easterly currents

The equatorial Atlantic Ocean surface circulation (Figure 1.14). It is possible to see the easterly South Equatorial Current that is straddling the Equator between 5S and 5N. The North equatorial Countercurrent is westerly and it covers the area between 5N and 10N, and the weaker North Equatorial Current is located at northern latitudes. The equatorial flow is divergent and also in this case we can presume the existence of upwelling at the Equator. A strong coastal current is visible as the Guinea Current, in the Gulf of Guinea. The South Equatorial Current feeds into the North Brazilian Current that then flows northward along the South America coast, taking different names as it finally emerges as the Caribbean Current in the Caribbean sea. A similar structure of alternating westerly and easterly undercurrents exist also in the equatorial Atlantic (Figure 1.14), but it is weaker and only the first westerly

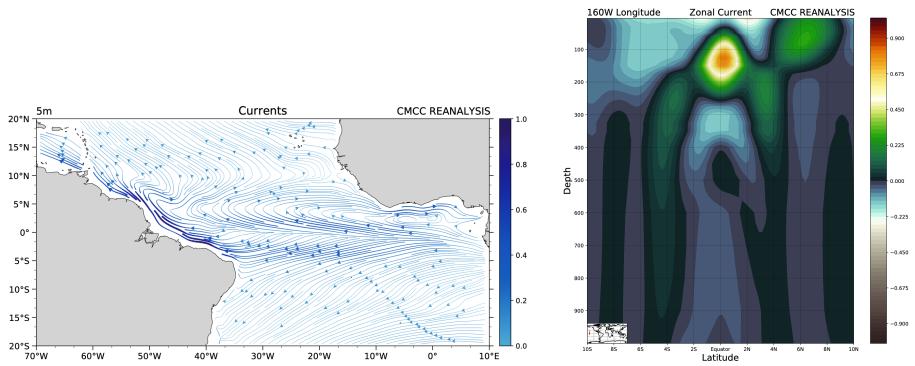


Figure 1.12: The equatorial Atlantic Ocean surface circulation and the structure of currents and undercurrents

maximum is well visible. It is also slanting towards the East, but the maximum is reached more towards the western boundary of the basin with respect to the Pacific Ocean. The deeper easterly jets are also much less weaker. The undercurrents jets are essentially absent in the Indian Ocean.

**The Gulf Stream** The current system of the Gulf Stream is one of the major feature of the global ocean circulation.

**The Kuroshio** The current system of the Kuroshio is one of the major features of the global ocean circulation.

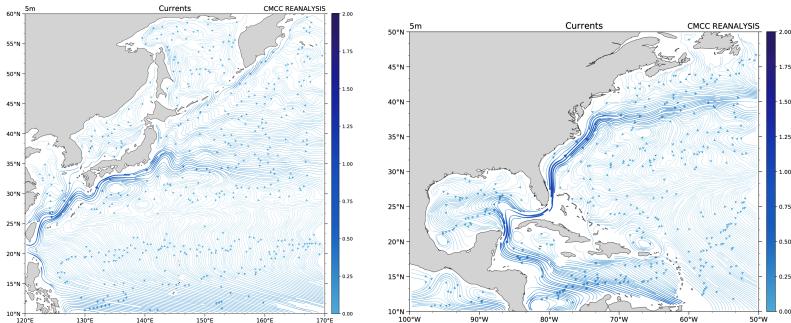


Figure 1.13: The Kuroshio on the left and Gulf Stream on the right

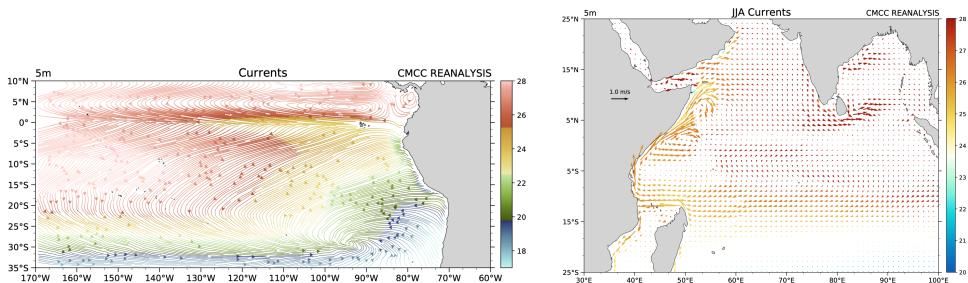


Figure 1.14: The upwelling zones on the left and the Somali current on the right.

### 1.3 Ice

Sea ice covers about 7% of Earth's oceans and plays a critical role in the global climate system, acting as a barrier between the atmosphere and the ocean. It forms when seawater cools below its freezing point, creating ice floes that are often cracked and shifted by environmental forces. Despite its compact appearance, sea ice is dynamic and never fully covers open water areas due to constant movement and deformation.

In the Southern Hemisphere, sea ice extends seasonally from about 3 million km<sup>2</sup> in summer to 18 million km<sup>2</sup> in winter, covering around 7% of the Southern Ocean at its peak. The ice can reach as far as 60S during winter, particularly in the Weddell and Ross Seas. However, it tends to be less compact compared to the Arctic, largely due to divergent winds and the influence of ocean currents. By contrast, in the Northern Hemisphere, sea ice ranges from 7 million km<sup>2</sup> in summer to 15 million km<sup>2</sup> in winter, making up about 10% of the ocean area during its maximum extent. The Arctic Ocean forms the core region of sea ice, with thick polar caps and thinner ice in surrounding seas like the Bering Sea and the Sea of Okhotsk. Ice can also drift into the North Atlantic through the Labrador and Greenland seas. The formation of sea ice depends on several factors. Salinity plays a crucial role, as higher salinity lowers the freezing point of seawater. In regions with salinity levels above 24.7%, convective mixing occurs as dense, salty water sinks, delaying the freezing process. Under calm conditions, supercooling can allow water to remain liquid below its freezing point until ice crystals form, which then accelerate the freezing of surrounding water. As the ice thickens, it insulates the ocean below, reducing heat loss and slowing further growth.

Sea ice melts during summer, particularly in the Arctic, where increased sunlight and prolonged daylight accelerate the process. Melting can occur at rates of about 40 mm per day, with melt ponds and drainage channels forming on the surface. The loss of ice is further aided by tidal forces and storms, leading to disintegration near coastlines and open water areas. By late summer, much of the peripheral sea ice in the Arctic disappears, leaving only the central ice pack.

Sea ice properties vary greatly between the Arctic and Antarctic regions. In the Arctic, central ice thickness typically ranges from 2 to 4 meters, with thinner ice in peripheral seas (0.5 to 1 meter). Antarctic icebergs are far larger, often exceeding 600 meters in thickness. Ice density, generally between 880 and 910 kg/m<sup>3</sup>, allows it to float. The topography of sea ice includes ridges and rubble zones formed by colliding ice sheets, with some ridges extending 10 to 15 meters below the surface.

The movement of sea ice is also an important factor in ocean circulation and climate. In the Arctic, the Beaufort Gyre and the Transpolar Drift Stream transport ice across the Arctic Basin, eventually exiting through the Fram Strait into the North Atlantic. This ice contributes to cold currents like the East Greenland Current. In the Antarctic, ice drifts clockwise in the Weddell Sea, eventually melting in warmer waters and transporting freshwater to lower latitudes.

Overall, sea ice is a vital component of Earth's climate system. It influences heat exchange between the ocean and atmosphere, regulates salinity and density in ocean waters, and contributes to the broader dynamics of polar and global ocean circulation. Its seasonal cycles and properties make it an essential focus for understanding climate change and its impacts.

### 1.4 Interconnections between atmosphere, ocean and ice

Important transfers of energy, mass, momentum occur between the oceans and the atmosphere and are greatly modified by the presence of sea ice.

### 1.4.1 Atmosphere effect

Wind forcing is a prime driver for upper ocean circulation and thereby impacts deep ocean circulation as well, and for sea ice thereby impacting its motions and location. In addition, air temperatures and moisture contribute to determining the energy fluxes across the interface between the atmosphere and surface, contributing to ice maintenance, growth and melt and to the temperature distribution in the ocean and water velocity. Fridtjof Nansen noted that sea ice does not drift in the direction of the wind but at an angle of 20 – 40 to the right due to the Earth's rotation and speculated that the motions in the water beneath sea ice deviate even more; the motion of each water layer deviate slower and farther to the right to the layer above and Ekman proved it mathematically adding that these influence stops at a depth of about 200 m.

### 1.4.2 Ocean effect

Ocean has a significant impact on the other two components. It is essentially a limitless moisture source and supplies the majority of atmospheric water vapor. The ocean land contrast exerts a major control over the distribution of evaporation over the global precipitation. The thermal inertia of the ocean results in a much lesser seasonal temperature range at the open ocean surface than at a land surface or a sea ice surface. This creates a lesser atmospheric temperature range over open ocean than over land or sea ice, as cold winter air is warmed by the underlying ocean and warm summer air is cooled. Since the ocean also absorbs a significant amount of the atmosphere's carbon dioxide ( $CO_2$ ), it also helps to slow the buildup of atmospheric  $CO_2$ .

The oceans contribute significantly to the poleward transport of heat, by means of ocean currents. This transport influences the heat balances in both high and low latitudes and the average temperature gradient between the equator and the poles. The temperature distribution of the ocean surface is the prime determinant of the sea ice distribution, because, in general, ice forms where the water temperature has reached the freezing point. Similarly, the ocean salinity distribution is important to the formation of sea ice, because the salt content of the water affects the freezing temperature. Once ice is formed, warm currents entering an ice-covered region tend to melt the ice cover, and cold currents moving away from the ice region tend to carry ice with them.

### **El Niño/Southern Oscillation (ENSO)**

This is an example of atmosphere/ocean interconnections one of the major large-scale patterns. The El Niño and Southern Oscillation phenomena were originally examined separately, before their close interconnection became apparent.

**El Niño** originally referred to a seasonal warming of the waters along the coast of Peru, frequently occurring shortly after Christmas. The term is now generally restricted to large-scale warming events, extending well into the central equatorial Pacific. The episodes can last many months, and do not necessarily begin near Christmas. **The Southern Oscillation**, characterized by consistent sea level pressure, temperature, and precipitation changes in the South Pacific, was first discussed by Walker who found an alternating pressure pattern involving the normal southeast Pacific high pressure and the low pressure region near the Indian Ocean and western Pacific regions.

*Under normal conditions* (non-El Niño), there is a strong sea surface temperature difference between the warm western Pacific and the cooler eastern Pacific, caused by upwelling along South America's west coast driven by east-to-west trade winds. This setup creates a convection cycle where air rises over the warm western Pacific, leading to heavy rainfall, and sinks over the cooler eastern Pacific, forming an east-west Walker circulation.

During *El Niño*, the trade winds weaken, reducing upwelling along the South American coast. This leads to warmer sea surface temperatures in the eastern Pacific, disrupting the temperature gradient. As a result, air rising and rainfall patterns shift eastward, cooling the western Pacific and causing significant climate effects. For example, the Indonesian-Australian region may experience drought, while wetter conditions occur farther east.

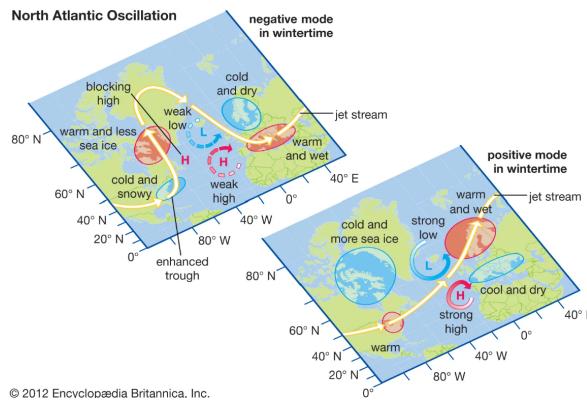


Figure 1.15: NAO

### North Atlantic Oscillation (NAO)

The North Atlantic Oscillation (NAO) is a pressure oscillation measured by the sea level pressure difference between Iceland and the Azores, reflecting the strength of the Icelandic low-pressure system and the Azores high-pressure system. (although sometimes it is indexed instead by the pressure difference between Newfoundland and Lisbon, Portugal, or between Iceland and Lisbon). It has both short-term and long-term variations, significantly affecting North Atlantic climate patterns. Positive NAO phases strengthen the Icelandic low pressure, resulting in colder winters in eastern North America, warmer and wetter conditions in western Europe, reduced sea ice near Greenland and Scandinavia, and increased sea ice in Baffin Bay and the Labrador Sea.

NAO impacts extend to global teleconnections, including influences on Russia, the Indian monsoon, and ocean temperatures. The positive NAO phase is associated with a tripolar sea surface temperature pattern: a cold anomaly in the subpolar North Atlantic, warm anomalies near Europe, and a cold subtropical anomaly near the Equator. The Gulf Stream further propagates these anomalies toward Europe, enabling atmospheric feedback and improving predictability of NAO patterns. The interaction between the ocean and the NAO remains a key area of research.

#### 1.4.3 Ice

The presence of sea ice has numerous climatic consequences, influencing the temperature and circulation patterns of both the atmosphere and the oceans. Sea ice lessens the amount of solar radiation absorbed at the ocean's surface: only about 20–50% of the incident solar radiation is absorbed, the rest being reflected to space and is therefore lost, while without the ice, typically 85–95% is absorbed. It serves as a strong insulator, restricting exchanges of heat from  $10^2$  to  $10^3 \text{ W m}^{-2}$  from the ocean to the atmosphere, to 10 to 20  $\text{W m}^{-2}$ ; mass, momentum, and chemical constituents between the ocean and atmosphere.

In winter, ice cover enhances polar cooling by increasing temperature gradients between the poles and the equator, intensifying atmospheric circulation. However, stronger circulation brings warm air into polar regions, reducing these gradients and creating a negative feedback, making the overall effect on atmospheric circulation uncertain. In summer, ice acts as a ther-

mal insulator, reducing heat transfer from the atmosphere to the ocean. Its high albedo also reflects solar radiation, limiting heat absorption and increasing the net heat gain in polar oceans. These seasonal dynamics highlight the complex role of ice in the climate system.

Another aspect of the insulation is the lessened evaporative transfer to the atmosphere in the presence of ice cover, resulting in reduction of moisture available for cloud formation, rain and snow.

The freezing and melting of sea ice influence seasonal and regional climates by moderating temperature extremes. Ice formation releases heat, while melting absorbs heat, facilitating a net equatorward transport of heat and salt from polar regions. During ice formation, salt is rejected into the underlying ocean, increasing the salinity and density of the mixed ocean layer, which can lead to deep convection and the formation of bottom water that drives global ocean circulation. This process is most significant at the edges of ice packs, where winds create open water that freezes rapidly, enhancing local ice production. Approximately one-third of bottom water originates from ice formation along ice margins. These dynamics connect ice processes to global climate systems, influencing both local and large-scale ocean circulation.

# Chapter 2

## Fundamental Equations and processes

Last updated: 2024-12-13

Source file: [chapter-fund-eqs-processes.tex](#)

### 2.1 Fundamental Equations

#### 2.1.1 Introduction

Sample citation of Richardson, [1922](#).

The atmosphere is in motion and it is a continuous mixing and clashing of vortices and structures, but when it is averaged over a long period of time (fig.2.1) it shows a remarkable simple structure. The figure shows the wind at an approximate level of about 12km, that is considered to be in the free atmosphere, far from the influence of the ground. The circulation is a large vortex around the pole that shows small oscillation in latitudes, especially pronounced over North America and the Asian Pacific Coast. The flow is therefore predominantly in the East-West direction, with a relatively small component in the meridional direction. The *circumpolar vortex* has been one of the first structures to be recognized when plentiful observations of the upper air flow became available, but it provided some of the intriguing questions that drove the development of geophysical fluid dynamics to this time, some of them have not been completely understood. What is maintaining this peculiar circulation? Which factor determines the amplitude and location of the undulations in meridional direction? Some of these question will be addressed in these chapter.

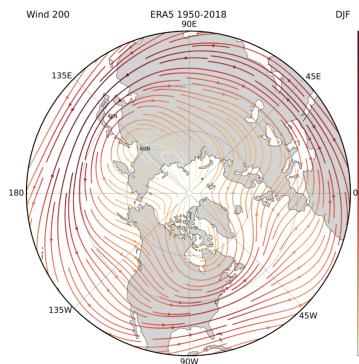


Figure 2.1: ERA5-wind 200

#### 2.1.2 Coordinate systems

**Spherical Coordinates** The most commonly used coordinate system for the analysis of the atmosphere and the oceans is a spherical coordinate system attached to the rotating Earth

(Fig. 2.2). The spherical coordinates are slightly different from the usual mathematical ones as the latitude is measured from the equator and therefore it can take negative values. The longitude is running west to east.

The longitude is also known as the "zonal" direction whereas the latitude is also known as the "meridional" direction. Winds are identified by the direction they are coming from, so a "westerly" wind is coming *from* the West and an "easterly" wind is coming from the East.

This coordinate system is rotating with the Earth and therefore it generates force terms in any dynamical equation expressed in this system of coordinate, the Coriolis terms.

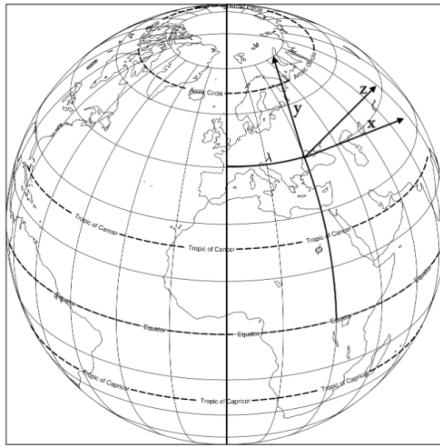


Figure 2.2: Coordinate system

### The Beta-plane

It is sometimes convenient to shift coordinate system if the latitudinal extension of the motion is not too great with respect to the motion parameters as they are expressed in the adimensional numbers. When this is possible, a tangent coordinate system is applied at a specific latitude  $\phi_0$  and the resultant Cartesian coordinates system is called the  $\beta$ -plane. Usually symbols  $(x, y)$  are used in this case for the zonal and meridional coordinate. In the  $\beta$ -plane the planetary vorticity  $f$  is linearized as  $f = f_0 + \beta y$ , where  $\beta = \frac{\partial f}{\partial y}(\phi_0)$ .

### Advection derivative

To describe the governing equation of the atmosphere and eventually of the ocean we have to understand how we write the rate of change with time of this fluid. This problem was solved by considering the fact that the rate of change in the fluid cannot be seen as rate of change with respect to a fixed system of coordinate because the system is moving with the fluid itself. Therefore, first we have to find a way to describe the change taking into account the moving system of restaurants. These can be done by using a concept developed in the 19th century by Euler that is called "advection derivative" that can be obtained from a total derivative of the property,

$$\frac{d\phi}{dt} = \frac{\partial\phi}{\partial t} + \frac{\partial x}{\partial t} \frac{\partial\phi}{\partial x} + \frac{\partial y}{\partial t} \frac{\partial\phi}{\partial y} + \frac{\partial z}{\partial t} \frac{\partial\phi}{\partial z} = \frac{\partial\phi}{\partial t} + \mathbf{v} \cdot \nabla \phi$$

and so it can be defined as

$$\frac{D\varphi}{Dt} = \frac{\partial\varphi}{\partial t} + \mathbf{v} \cdot \nabla \varphi$$

in this way the moving fluid can be described by derivatives with respect the "fixed" coordinate system, i.e. the Eulerian description. The alternative description of the observer moving with fluid is known as the "Lagrangian" description.

### 2.1.3 Equation of motion

The motion of a physical system is governed by conservation laws: conservation of momentum, mass and energy. The conservation of momentum equation will identify the forces acting on the system. The conservation of energy will identify the processes capable of changing the energy of the system, i.e. thermodynamical processes. The equation governing the motion of the atmosphere can be written as:

$$\begin{aligned}\frac{Du}{Dt} - \frac{uv \tan \phi}{r} + \frac{uw}{r} &= -\frac{1}{\rho r \cos \phi} \frac{\partial p}{\partial \lambda} + fv - \hat{f}w + F_\lambda \\ \frac{Dv}{Dt} - \frac{u^2 \tan \phi}{r} + \frac{vw}{r} &= -\frac{1}{\rho r} \frac{\partial p}{\partial \phi} - fu + F_\phi \\ \frac{Dw}{Dt} - \frac{u^2 + v^2}{r} &= -\frac{1}{\rho} \frac{\partial p}{\partial z} - g + \hat{f}u + F_z\end{aligned}$$

the  $f = 2\Omega \sin \phi$  and  $\hat{f} = 2\Omega \cos \phi$  terms arise from the rotating spherical coordinate system that we have chosen, other terms are generated by the spherical geometry. Some of them are small and traditionally they can be neglected, so that we arrive at the system

$$\begin{aligned}\frac{Du}{Dt} - v \left( f + \frac{u \tan \phi}{a} \right) &= -\frac{1}{\rho a \cos \phi} \frac{\partial p}{\partial \lambda} + F_\lambda \\ \frac{Dv}{Dt} + u \left( f + \frac{u \tan \phi}{a} \right) &= -\frac{1}{\rho a} \frac{\partial p}{\partial \phi} + F_\phi \\ \frac{Dw}{Dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial z} - g + F_z\end{aligned}$$

where we have also used the *Shallowness Approximation* by assuming  $r = a + z \approx a$ , where  $a$  is the Earth radius.

However the advective derivative must be expressed in spherical coordinates

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \frac{u}{a \cos \phi} \frac{\partial}{\partial \lambda} + \frac{v}{a} \frac{\partial}{\partial \phi} + w \frac{\partial}{\partial z}$$

so that the velocity components are

$$\begin{aligned}u &= a \cos \phi \frac{\partial \lambda}{\partial t} \\ v &= a \frac{\partial \phi}{\partial t} \\ w &= \frac{\partial z}{\partial t}\end{aligned}$$

These equations govern the mechanical behaviour of the atmosphere, and we will see in a different form, also of the ocean. There three forces in action: pressure gradient, rotation via the Coriolis force and gravity.

The equation are not complete, we have three equation but five variables, so we need to find the missing relations. We are using the basic conservation principles, the latter equations describe the conservation of momentum, we can exploit the conservation of mass. The mass of the fluid must be conserved locally, because there are now sinks or sources in the atmosphere itself, so we want to write the mass of a volume of atmosphere fixed in space as

$$M = \int_V \rho dV$$

the mass in the volume can only change if there is a flux of mass at surface  $S$ ,

$$\frac{\partial}{\partial t} \int_V \rho dV = - \int_S \rho \mathbf{v} \cdot \mathbf{n} dS$$

using the divergence theorem however we have

$$\frac{\partial}{\partial t} \int_V \rho dV = - \int_V \nabla \cdot (\rho \mathbf{v}) dV$$

because the volume is not changing with time we can bring the derivative inside the integral and we get

$$\int_V \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) dV = 0$$

but the volume is arbitrary, so it must be that

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

is valid locally.

We have still at our disposal the conservation of thermodynamical energy and so we can also use the first law of thermodynamics for a gas, that is a statement of internal energy, where  $C_v$  is the specific heat for air at constant volume and  $T$  is the temperature in Kelvins.

$$c_v \frac{DT}{Dt} = -p \frac{D}{Dt} \left( \frac{1}{\rho} \right) + Q$$

where we included the temperature and heating/cooling term  $Q$  (which is the net heat gain or loss to the external sources, for example the solar insolation, heating or cooling due to long wave radiation, latent heating due to condensation of water vapor into liquid water, and sensible heating due to conduction and convection). The state variable are then linked by the state equation

$$p = \rho R T$$

where  $R$  is the gas constant for dry air.

We can use the equation of state to write the energy equation (or the temperature equation) in a different form,

$$c_v \frac{DT}{Dt} = -p \frac{D}{Dt} \left( \frac{RT}{p} \right) + Q = -R \frac{DT}{Dt} + \frac{RT}{p} \frac{Dp}{Dt} + Q$$

yielding the alternative forms ( since  $c_p = c_v + R$ ),

$$c_p \frac{DT}{Dt} - \frac{1}{\rho} \frac{Dp}{Dt} = Q$$

For many purpose atmospheric and oceanic motions can be considered essentially adiabatic. However, for climate studies, the assumption of exclusively adiabatic processes is not appropriate since the amount of heat added or lost to a unit volume of air or water over a long period of time can be substantial. For adiabatic processes  $Q = 0$ :

$$\begin{aligned} c_p \frac{DT}{Dt} - \frac{1}{\rho} \frac{Dp}{Dt} &= 0 \\ \frac{c_p}{T} \frac{DT}{Dt} - \frac{R}{p} \frac{Dp}{Dt} &= 0 \\ \frac{D}{Dt} \log T - \frac{R}{c_p} \frac{D}{Dt} \log p &= 0 \end{aligned}$$

integrating it we get

$$\log T/T_0 - \log \left( \frac{p}{p_0} \right)^{R/c_p} = \text{const}$$

or

$$\frac{T}{T_0} \left( \frac{p_0}{p} \right)^{R/c_p} = \text{const}$$

so the quantity, known as *potential temperature*

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

is conserved in adiabatic processes and the thermodynamics equation can be written as

$$\frac{D\theta}{Dt} = Q$$

**Hydrostatic balance.** Under the action of gravity the vertical component of the pressure gradient force balances the action of gravity, resulting in very small vertical acceleration

$$\frac{\partial p}{\partial z} = -g\rho$$

then if we take the vertical derivative of the eq. Eq:logT

$$\frac{1}{T_0} \frac{dT}{dz} \left( \frac{p_0}{p} \right)^{R/c_p} - \frac{p_0}{p^2} \frac{R}{c_p} \frac{T}{T_0} \left( \frac{p_0}{p} \right)^{R/c_p-1} \frac{dp}{dz} = 0$$

simplifying

$$\frac{dT}{dz} - \frac{p_0}{p^2} \frac{R}{c_p} T \left( \frac{p_0}{p} \right)^{-1} \frac{dp}{dz} = 0$$

or

$$\frac{dT}{dz} - \frac{1}{p} \frac{R}{c_p} T \frac{dp}{dz} = \frac{dT}{dz} + g\rho \frac{1}{p} \frac{R}{c_p} T = 0$$

but using the equation of state

$$\frac{dT}{dz} = -\frac{g}{c_p}$$

that gives how the temperature change with height under adiabatic conditions and when the hydrostatic balance is valid. This is known as the *adiabatic lapse rate*. If I assume:

$$C_p = 1000; \frac{-g}{C_p} = 10C/km$$

the temperature must drops linearly with height.

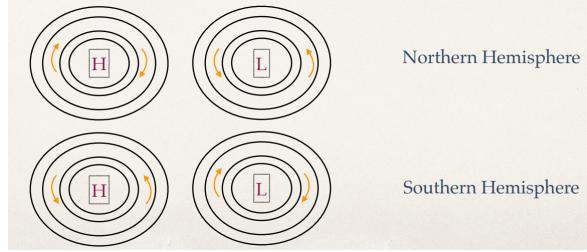


Figure 2.3: Geostrophic balance

If we integrate over height  $z$ , assuming isothermal atmosphere, we get an exponential trend over density  $\rho$  and pressure  $p$ , which drops with height (because of the perfect gas law):

$$p(z) = p_0 e^{\frac{-RT}{g} z}$$

**Geostrophic balance.** Geostrophic balance is the balance between Coriolis Force and the horizontal component of the pressure gradient force. The pressure gradient force (PGF) pushes air or water from high pressure to low pressure. The Coriolis force deflects the motion to the right in the Northern Hemisphere and to the left in the Southern Hemisphere. In geostrophic balance, these forces are equal in magnitude and opposite in direction. The geostrophic balance assumes the flow is steady and does not accelerate, so it neglects the effects of inertia. This balance is most accurate for large-scale motions (e.g., planetary or synoptic scales) where friction and other forces are negligible. Vertical motions are typically much smaller than horizontal motions and are neglected in geostrophic balance.

In geostrophic balance:

- Air or water flows parallel to the isobars (lines of constant pressure) or contours of constant geopotential height.
- The speed of the geostrophic flow increases with stronger pressure gradients (closer isobars).

However, on small scales (e.g., tornadoes, boundary layers), friction and other forces become important, so geostrophic balance is less accurate. Near the Equator, The Coriolis parameter ( $f$ ) approaches zero, making the geostrophic balance invalid.

**Advection derivative in rotating coordinates.** A vector in spherical coordinates is:

$$\mathbf{u} = \hat{i}u + \hat{j}v + \hat{k}w$$

but the unit vectors move in a rotating frame, so the advective derivative of a vector is given by:

$$\frac{D\mathbf{u}}{Dt} = \frac{Du}{Dt}\hat{i} + \frac{Dv}{Dt}\hat{j} + \frac{Dw}{Dt}\hat{k} + u\frac{D\hat{i}}{Dt} + v\frac{D\hat{j}}{Dt} + w\frac{D\hat{k}}{Dt} \quad (2.1)$$

**Coriolis Factors** The Coriolis forces derives from the conservation of angular momentum.

$$\frac{d}{dt}(R_A(\Omega R_A + u)) = 0$$

or

$$2\Omega R_A \frac{dR_A}{dt} + u \frac{dR_A}{dt} + R_A \frac{du}{dt} = 0$$

rearranging

$$\frac{du}{dt} = (2\Omega R_A + u) \frac{dR_A}{dt}$$

The Coriolis components then are  $2\Omega \sin \phi$  and  $2\Omega \cos \phi$ . For the horizontal and vertical components respectively:  $f = 2\Omega \sin \phi$ .

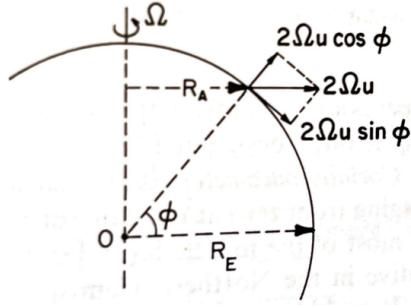


Figure 2.4: Coriolis factors

#### 2.1.4 Summary of fundamental equations-

Summarizing our discussion, the fundamental equation that describe the motion of the atmosphere are considered in the following approximations:

- Hydrostatic approximation
- Shallow fluid approximation. The vertical coordinate  $r$  is substituted by  $a + z(a \gg z)$  except in differentiation
- Neglecting metric terms involving vertical velocity  $w$

The first approximation is independent, the second and the third must be applied together. These primitive (non hydrostatic) equations are:

$$\begin{aligned} \frac{Du}{Dt} - v \left( f + \frac{u \tan \phi}{a} \right) &= -\frac{1}{a \cos \phi} \frac{1}{\rho} \frac{\partial p}{\partial \lambda} + F_\lambda \\ \frac{Dv}{Dt} + u \left( f + \frac{u \tan \phi}{a} \right) &= -\frac{1}{a \rho} \frac{1}{\partial \phi} \frac{\partial p}{\partial \phi} + F_\phi \\ \frac{Dw}{Dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial z} - g + F_z \\ \frac{D\theta}{Dt} &= Q \\ \frac{\partial \rho}{\partial t} + \frac{1}{a \cos \phi} \left[ \frac{\partial}{\partial \lambda} (\rho u) + \frac{\partial}{\partial \phi} (rv \cos \phi) \right] + \frac{\partial}{\partial z} (\rho w) &= 0 \\ p &= \rho RT \end{aligned}$$

where we have used the divergence in spherical coordinates.

These equations are still not closed because we will need to express the heating/cooling term  $Q$  and the friction terms  $F$  as a function of the state variables. This will require a theory of the processes that drive them. Where  $R = 287.052874 J \text{ kg}^{-1} \text{ K}^{-1}$  is the gas constant for dry air and  $c_p = 1.005$  is the specific heat at constant pressure,  $c_v = 0.718$  is the specific heat at constant volume,  $\kappa = \frac{R}{c_p}$  and  $\gamma = c_p/c_v$  is their ratio.

#### 2.1.5 Simplified equations-

For theoretical and idealized studies the set of equation projected on the  $\beta$ -plane is also used. The  $\beta$ -plane approximation is a simplified model used in geophysical fluid dynamics to account the variation of the Coriolis parameter  $f$  with latitude, in this approximation  $f$  is linearized around a reference latitude such that  $f \approx f_0 + \beta y$  with  $y = a(\phi - \phi_0)$  is the northward displacement from the reference latitude. With the  $\beta$ -plane approximation, no rotation, no sphericity, neglecting the meridional coordinates , the equations become:

$$\begin{aligned}\frac{Du}{Dt} - fv &= -\frac{1}{\rho} \frac{\partial p}{\partial x} + F_x \\ \frac{Dv}{Dt} + fu &= -\frac{1}{\rho} \frac{\partial p}{\partial y} + F_y \\ \frac{Dw}{Dt} &= -\frac{1}{\rho} \frac{\partial p}{\partial z} - g + F_z \\ \frac{D\theta}{Dt} &= Q \\ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) &= 0 \\ p &= \rho RT\end{aligned}$$

and the gradient operator is the cartesian operator

$$\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$$

and the advective derivative is then

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + w \frac{\partial}{\partial z}$$

The first set of fundamental equations that we encounter is the conservation of Momentum (**Navier-Stokes equations**).

The Navier-Stokes equations describe the motion of fluid substances like air and water. They account for forces due to pressure, viscosity, and external forces such as gravity. This set of equations is key to understanding how winds, ocean currents, and other flows evolve due to internal and external forces. In a rotating frame of reference, specifically describing the motion of a fluid (e.g. the atmosphere or ocean currents) the principal equation can be declined in the different directions: longitude, latitude and vertical directions.

The left-hand side of the equations begins with material derivative of the velocity components ( $u$  in the east-west direction, associated with longitude;  $v$  north-south velocity component, associated with latitude,  $w$  is the vertical velocity component). The material derivative includes both the local rate of change and the advection (transport) of the velocity. This term tells us how the velocity of a fluid parcel changes over time, taking into account both temporal changes and the movement of the parcel.

The second term on the left represents the Coriolis force and the centrifugal force in the rotating reference frame of the Earth.  $f$  is the Coriolis parameter, which depends on the latitude  $\phi$  and is given by  $f = 2\Omega \sin \phi$  where  $\Omega$  is the angular velocity of the Earth and  $\phi$  is the latitude. The Coriolis force is proportional to the velocity and acts perpendicular to the motion of the fluid, deflecting the fluid in different directions depending on the hemisphere.

On the right side of the equations we find the pressure gradient force in the longitudinal (first) and latitudinal (second) directions, which drives fluid motion due to differences in pressure. The terms  $F_\lambda$ ,  $F_\phi$  and  $F_z$  represent an additional force term that could represent any other forces acting on the fluid parcel in the longitudinal, latitudinal and vertical direction. It might include friction, external forces, or any other model-specific forces not accounted for in the other terms.

Let's now focus on them one by one.

$$\frac{Du}{Dt} - v \left( f + \frac{u \tan \phi}{a} \right) = -\frac{1}{a \cos \phi} \frac{1}{\rho} \frac{\partial p}{\partial \lambda} + F_\lambda \quad (2.2)$$

The term  $\frac{u \tan \phi}{a}$  accounts for the centrifugal force resulting from the Earth's rotation. Here,  $u$  is the east-west velocity,  $\phi$  is the latitude, and  $a$  is the Earth's radius. The centrifugal force

is stronger near the equator, and this term adjusts the Coriolis effect to account for that.  $\frac{\partial p}{\partial \lambda}$  is the derivative of pressure with respect to longitude ( $\lambda$ ), indicating how pressure changes as you move east or west. The term  $a \cos \phi$  accounts for the spherical geometry of the Earth and the fact that distances between lines of longitude vary with latitude (they are widest at the Equator and shrink towards the poles). This term describes the acceleration due to the horizontal pressure gradient in the longitudinal direction, with the pressure gradient force causing flow from regions of higher to lower pressure.

$$\frac{Dv}{Dt} + u \left( f + \frac{v \tan \phi}{a} \right) = -\frac{1}{a \rho} \frac{\partial p}{\partial \phi} + F_\phi \quad (2.3)$$

The term  $\frac{v \tan \phi}{a}$  is the centrifugal force term in the north-south direction, considering Earth's curvature.  $\frac{p}{\phi}$  represents the pressure gradient in the latitude direction, and it drives motion from high to low pressure.

Note that the Coriolis term arises due to the rotation of the Earth, which introduces an apparent force in a rotating reference frame. This term depends explicitly on the latitude because of how the Earth's rotation affects the direction and magnitude of the Coriolis effect at different points on the Earth's surface. The Coriolis force explicitly depends on the sine of the latitude, as the projection of  $\Omega$  onto the local horizontal plane is  $\Omega \sin \phi$  (the latitude  $\phi$  determines the angle between the Earth's rotational axis and the local vertical).  $\Omega$  is the angular velocity vector of Earth's rotation, it points along the axis of rotation (towards the North Pole). At the Equator  $\phi = 0$  the Coriolis effect is perpendicular to  $\Omega$  and has a max horizontal effect. At the poles ( $\phi = \pm 90$ ) the effect aligns with  $\Omega$ , and only the vertical motion is affected.

In the **vertical direction**:

$$\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g + F_z \quad (2.4)$$

While the Coriolis force primarily affects horizontal motion, the vertical component of the Coriolis force is negligible because  $\Omega$  is nearly parallel to the vertical axis in most regions. In rotating systems like the atmosphere or oceans, vertical Coriolis terms are often ignored.  $g$  acts downward and opposes upward motion.

The equation

$$\frac{D\theta}{Dt} = Q \quad (2.5)$$

says that the rate of change of  $\theta$  (the material derivative) experienced by a fluid parcel moving through space is equal to the source term  $Q$ . This equation describes the rate of change of a scalar quantity (like temperature or moisture) **for a fluid parcel** moving through space. The term  $Q$  could represent a source of energy or mass, such as heat from the sun or moisture added by evaporation. In the context of a weather or climate model, this would represent how temperature (or another variable) changes as the air parcel moves and as energy is gained or lost. This equation is a general form for describing the time evolution of a scalar quantity (like T or moisture) in a moving fluid, where the change in that quantity is driven by external sources or sinks.

The **conservation of mass** or **continuity equation** ensures that the mass is conserved in the system. For the atmosphere or the ocean, it expresses how the density of air or water changes over time due to processes like flow and diffusion.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (2.6)$$

$\nabla \cdot (\rho \mathbf{v})$  represents the divergence of the mass flux, which accounts for the movement of the mass.

The last one is the equation of state for atmospheric or oceanic models links pressure, density and temperature of the fluid. The **ideal gas law** is typically used:

$$p = \rho RT \quad (2.7)$$

### 2.1.6 Linear solutions

We look for solution that describe small oscillation away from a basic state, in this case assumed in hydrostatic balance with constant wind of velocity  $u$ , and I'm looking at deviation from this basic state:

$$\frac{\partial p_0}{\partial z} = -g\rho_0 = -\frac{gp_0}{RT_0} \quad (2.8)$$

With the further assumption of isotherm atmosphere we can integrate to get:

$$p_0(z) = p_R e^{-z/H} \quad (2.9)$$

and  $H = \frac{RT_0}{g}$  is the scale height.

As we have seen, the primitive equations are a system of nonlinear partial differential equations governing fluid motion on a rotating sphere. These include momentum equations, continuity (mass conservation), thermodynamic equations and the equation of state. Since solving the full nonlinear system is often impractical, the system is linearized to analyze small deviations (or perturbations) from a reference state (steady flow or a state of rest). Assume the flow can be separated into a basic state and a perturbation:

$$\begin{aligned} u &= u_0 + u' \\ v &= v_0 + v' \\ \theta &= \theta_0 + \theta' \\ p &= p_0 + p' \end{aligned}$$

The resulting linearized primitive equations describe how perturbations evolve over time and space (linear equations are obtained for the prime variables neglecting all terms quadratic in perturbation). For instance, the pressure gradient becomes:

$$\frac{1}{\rho_0 + \rho'} \nabla(p_0 + p') + g\hat{k} \approx \frac{1}{\rho_0 + \rho'} \nabla p' + g\hat{k} = \frac{1}{\rho_0} \frac{1}{1 + \frac{\rho'}{\rho_0}} \nabla p' + g\hat{k} = \frac{1}{\rho_0} \nabla p' + g\hat{k} \quad (2.10)$$

and the potential temperature:

$$\theta = T \left( \frac{p_R}{p} \right)^k = \frac{p_R^k}{R} \frac{p^{1-k}}{\rho} \approx \theta_0 \left[ 1 + \frac{1}{\gamma} \left( \frac{p'}{p_0} \right) - \frac{\rho'}{\rho_0} \right]$$

hence,

$$\frac{\theta'}{\theta_0} = \frac{1}{\gamma} \frac{p'}{p_0} - \frac{\rho'}{\rho_0} \quad (2.11)$$

with  $\gamma = \frac{c_p}{c_v}$ ,  $k = \frac{R}{c_v}$ ,  $R = c_p - c_v$ .

### 2.1.7 Waves

The linearized system often supports wave-like solutions. These solutions arise naturally due to the restoring forces in the equations, such as:

- Pressure gradient drives the oscillations through compressibility or buoyancy
- Coriolis force introduces rotational effects leading to inertial and planetary waves
- Gravity acts as a restoring force for vertical displacement, driving internal gravity waves.

We take the equations for perturbation:

$$\begin{aligned}\frac{\partial u}{\partial t} + U \frac{\partial u}{\partial x} + \frac{1}{\rho_0} \frac{\partial p}{\partial x} &= 0 \\ \frac{\partial w}{\partial t} + U \frac{\partial w}{\partial x} + \frac{1}{\rho_0} \frac{\partial p}{\partial z} + g \frac{\rho}{\rho_0} &= 0 \\ \frac{\partial \rho}{\partial t} + U \frac{\partial \rho}{\partial x} + w \frac{\partial \rho_0}{\partial z} + \rho_0 \left( \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} \right) &= 0 \\ \frac{\partial \theta}{\partial t} + U \frac{\partial \theta}{\partial x} + w \frac{\partial \theta_0}{\partial z} &= 0 \\ \frac{\theta}{\theta_0} + \frac{\rho}{\rho_0} + \frac{1}{\gamma} \frac{p}{p_0} &= 0\end{aligned}$$

Basically to find solutions, assume the perturbations take the form of plane waves:

$$\mathbf{u}'(x, t) = \hat{\mathbf{u}} e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}$$

where  $\hat{\mathbf{u}}$  is the amplitude of the perturbation,  $\mathbf{k}$  is the wavevector,  $\omega$  the angular velocity. Substituting this kind of wave into the linearized equations leads to a dispersion relation, which relates  $\omega$  to  $\mathbf{k}$  and the physical parameters of the system (Coriolis, ...). The perturbation equations link the restoring forces to wave properties:

1. The momentum equations describe how velocity perturbations interact with pressure gradients and Coriolis forces.
2. The continuity equation ensures mass conservation, connecting velocity and density perturbations.
3. The thermodynamic equation relates temperature, pressure and density perturbations

Imposing the hydrostatic approximation from the beginning would mean putting to zero from the start the time derivative of  $w$  in the momentum equation in the vertical. This basically is equivalent to say that we are removing the explicit dependence of the vertical velocity on time and therefore we will get waves similar to the Boussinesq approximation. The hydrostatic approximation eliminates vertical propagating sound waves, but the Lamb waves still exists.

## 2.2 Homogeneous flows

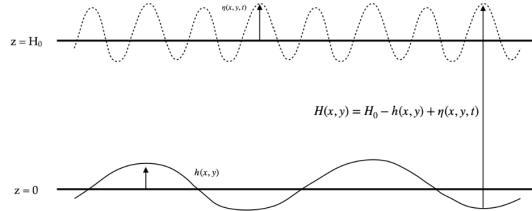


Figure 2.5: Homogeneous flow

The motion of the atmosphere is most appropriately described by three-dimensional equations that describes the horizontal and vertical motion of the fluid, however a lot can be

understood by considering a simpler system that consider the motion of a free surface of a inviscid, homogeneous and incompressible fluid. These equations are variously referred to as one-level primitive equations or the *shallow water equations*.

The system is described in Fig. 2.5, it is an homogenous layer of fluid covering the entire spherical planet. We have included a bottom topography  $h(x, y)$  and a deformation of the free surface  $\eta(x, y)$ . The convention is such positive deformation of the surface are increasing the depth of the fluid, whereas positive deformation of the bottom are decreasing it.

We have introduced here cartesian coordinates  $(x, y)$  such that  $dx \approx a \cos \theta d\lambda$  and  $dy \approx a d\theta$ , we will also define  $f = 2\Omega \sin \theta$ , then the incompressible equations of motion can be written as

$$\begin{aligned}\frac{\partial u}{\partial t} &= -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - \frac{1}{\rho} \frac{\partial p}{\partial x} \\ \frac{\partial v}{\partial t} &= -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - \frac{1}{\rho} \frac{\partial p}{\partial y} \\ \frac{\partial w}{\partial t} &= -u \frac{\partial w}{\partial x} - v \frac{\partial w}{\partial y} + g - \frac{1}{\rho} \frac{\partial p}{\partial z} \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} &= 0\end{aligned}$$

If the aspect ratio  $\delta \approx H/L$  is small then hydrostatic balance is maintained to order  $\delta^2$  and so

$$\frac{\partial p}{\partial z} = -\rho g + O(\delta^2)$$

because the density is constant we can integrate it from 0 to  $z$  and get

$$p = -\rho g (H_0 + \eta - z) + p_0$$

where we have used the boundary condition at the top  $p(x, y, H_0 + \eta) = p_0$ .

The horizontal pressure gradients are independent of  $z$

$$\begin{aligned}\frac{\partial p}{\partial x} &= \rho g \frac{\partial \eta}{\partial x} \\ \frac{\partial p}{\partial y} &= \rho g \frac{\partial \eta}{\partial y}\end{aligned}$$

because the horizontal accelerations are independent of  $z$ , than also the velocities are independent of  $z$ , if they are so at the beginning. This is in fact a consequence of the Taylor-Proudman theorem applied to a homogeneous fluid. We can then write the horizontal momentum equations

$$\begin{aligned}\frac{\partial u}{\partial t} &= -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - g \frac{\partial \eta}{\partial x} \\ \frac{\partial v}{\partial t} &= -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - g \frac{\partial \eta}{\partial y}\end{aligned}$$

Because  $u$  and  $v$  are independent of  $z$  allows us to integrate vertically the divergence equation from the surface  $h(x, y)$  to an height  $z$

$$w(x, y, z, t) = -z \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \omega(x, y, t)$$

The kinematic condition for the bottom can be written as

$$\frac{D}{Dt}(z - h) = 0$$

or

$$w(x, y, h, t) = u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y}$$

so the function  $\omega$  is

$$\omega(x, y, t) = u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y} + h \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)$$

then the vertical velocity can be written as

$$w(x, y, z, t) = (h - z) \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y}$$

Using now the kinematic boundary condition at the top ( $z = H_0 + \eta$ )

$$w(x, y, H_0 + \eta, t) = \left( \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} \right) (H_0 + \eta)$$

combining the last two equation we get an equation for the height

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x} (H_0 + \eta - h) u + \frac{\partial}{\partial y} (H_0 + \eta - h) v = 0$$

and so we can now write the complete equations, introducing the total depth of the fluid  $H = H_0 + \eta - h$ :

$$\begin{aligned} \frac{\partial u}{\partial t} &= -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - g \frac{\partial H}{\partial x} \\ \frac{\partial v}{\partial t} &= -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - g \frac{\partial H}{\partial y} \\ \frac{\partial H}{\partial t} &= -\frac{\partial(uH)}{\partial x} - \frac{\partial(vH)}{\partial y} \end{aligned}$$

These equations are written in the approximation of no friction, molecular or turbulent;  $w \ll u$ , from a scale analysis of

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = -\frac{\partial w}{\partial z}$$

Linearizing around a state of rest, constant rotation:

$$\begin{aligned} \frac{\partial u}{\partial t} - fv &= -g \frac{\partial \eta}{\partial x} \\ \frac{\partial v}{\partial t} + fv &= -g \frac{\partial \eta}{\partial y} \\ \frac{\partial \eta}{\partial t} &= -H_0 \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \end{aligned}$$

the solutions are, respectively:

$$\begin{aligned} u(x, y, t) &= \operatorname{Re} \left[ \tilde{u} \sin(l y) e^{i(kx - \omega t)} \right] \\ v(x, y, t) &= \operatorname{Re} \left[ \tilde{v} \sin(l y) e^{i(kx - \omega t)} \right] \\ \eta(x, y, t) &= \operatorname{Re} \left[ \tilde{\eta} \sin(l y) e^{i(kx - \omega t)} \right] \end{aligned}$$

two modes<sup>1</sup>, a stationary mode and a gravity mode (Poincare' modes):  $\omega = 0$  and  $\omega^2 = f^2 + gH_0(k^2 + l^2)$

---

<sup>1</sup>mode in this context refers to a specific solution or pattern of motion that satisfies the governing equations of the flow, under certain boundaries or initial conditions.

### 2.2.1 The vorticity equation on the $\beta$ -plane

The spherical geometry maybe cumbersome without adding much the conceptual discussions, so it may be convenient to introduce Cartesian coordinates  $(x, y)$  such that  $dx \approx a \cos \theta d\lambda$  and  $dy \approx a d\theta$ , we will also define  $f = 2\Omega \sin \theta$ , then we obtain equation formally identical to the Cartesian equation described previously

$$\begin{aligned}\frac{\partial u}{\partial t} &= -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - g \frac{\partial \eta}{\partial x} \\ \frac{\partial v}{\partial t} &= -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - g \frac{\partial \eta}{\partial y} \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= 0\end{aligned}$$

where we have kept the pressure terms, though they are zero in this case ( $h$  is a constant) as a placeholder. In vector notation (all vectors are two-dimensional)

$$\frac{\partial \mathbf{v}}{\partial t} = -(\mathbf{v} \cdot \nabla) \mathbf{v} - f(\hat{k} \times \mathbf{v}) - \nabla \eta$$

using the vector identity<sup>2</sup>

$$\frac{1}{2} \nabla(\mathbf{v} \cdot \mathbf{v}) = (\mathbf{v} \cdot \nabla) \mathbf{v} + \mathbf{v} \times (\nabla \times \mathbf{v})$$

so substituting in the equation

$$\frac{\partial \mathbf{v}}{\partial t} = -(\zeta + f)\hat{k} \times \mathbf{v} - \nabla(\eta + \frac{1}{2}|\mathbf{v}^2|) \quad (2.12)$$

(for 2-dimensional flows  $\nabla \times \mathbf{v} = \zeta \hat{k}$ ).

and to express more clearly the components

$$\begin{aligned}\frac{\partial u}{\partial t} &= (\zeta + f)v - \frac{\partial}{\partial x}(\eta + \frac{1}{2}|\mathbf{v}^2|) \\ \frac{\partial v}{\partial t} &= -(\zeta + f)u - \frac{\partial}{\partial y}(\eta + \frac{1}{2}|\mathbf{v}^2|)\end{aligned}$$

By taking the curl of eq.2.12 we obtain an equation for the vorticity (as it will be discussed in Sec.3.8):

$$\begin{aligned}\frac{\partial \zeta}{\partial t} &= -\frac{\partial}{\partial x}(\zeta + f)u - \frac{\partial}{\partial y}(\zeta + f)v \\ &\quad - \frac{\partial}{\partial x} \frac{\partial}{\partial y}(p + \frac{1}{2}|\mathbf{v}^2|) + \frac{\partial}{\partial y} \frac{\partial}{\partial x}(p + \frac{1}{2}|\mathbf{v}^2|) \\ &= -\frac{\partial}{\partial x}(\zeta + f)u - \frac{\partial}{\partial y}(\zeta + f)v = -\mathbf{v} \cdot \nabla(\zeta + f)\end{aligned}$$

where we have used the non divergence condition  $\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$  in the last step. Once again, this is the equation for the conservation of total vorticity where we can see that the relative vorticity is a sort of additional Coriolis effect (or alternatively that the Coriolis term is a source of vorticity). Because of the  $\beta$ -plane  $f = f_0 + \beta y$ :

$$\frac{\partial \zeta}{\partial t} = -u \frac{\partial \zeta}{\partial x} - v \frac{\partial \zeta}{\partial y} - \beta v \quad (2.13)$$

---

<sup>2</sup>For a general vector  $\mathbf{A}$ :

$$\frac{1}{2} \nabla(A \cdot A) = (A \cdot \nabla)A + A \times (\nabla \times A)$$

that is the non-divergent barotropic vorticity equation. And the potential vorticity:

$$\frac{D}{Dt}(\zeta + f) = 0 \quad (2.14)$$

$f = 2\Omega \sin(\theta)$  is the Coriolis coefficient, that is the vertical component of the Earth planetary vorticity.

In the new coordinates  $(x, y)$  for longitude and latitude, the streamfunction and the vorticity then are

$$u = -\frac{\partial \psi}{\partial y} \quad v = \frac{\partial \psi}{\partial x} \quad \zeta = \frac{\partial x}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2 \psi$$

introducing the Jacobian operator

$$J(A, B) = \frac{\partial A}{\partial x} \frac{\partial B}{\partial y} - \frac{\partial A}{\partial y} \frac{\partial B}{\partial x}$$

we obtain a single equation for the streamfunction

$$\frac{\partial}{\partial t} \nabla^2 \psi + J(\psi, \nabla^2 \psi + f_0 + \beta y) = 0$$

that represents the quasi-geostrophic potential vorticity. Vorticity is a vertical component of the velocity, and appears coupled with  $f$ , where  $f$  is a property of the Earth and  $\zeta$  is a property of the flow. Because the potential vorticity is conserved with the flow, if you have something at rest at the Equator and you move it Poleward, it acquire vorticity and starts speeding .

## 2.3 Rossby waves

Rossby waves or planetary waves are large-scale waves that arise in a rotating fluid, such as the Earth's atmosphere or oceans, due to the variation of the Coriolis parameter  $f$  with latitude. In particular, they are caused by the  $\beta$ -effect, i.e.  $\beta = \frac{\partial f}{\partial y} = \frac{2\Omega \cos \phi}{a}$ . These waves typically have very large horizontal wavelengths and govern large-scale motions in the atmosphere and oceans, their restoring force is the variation of the Coriolis force with latitude. Rossby waves are much slower than other atmospheric or oceanic waves, such as gravity or sound waves.

Linearizing around a basic state  $U$ :

$$\frac{\partial \zeta'}{\partial t} + U \frac{\partial \zeta'}{\partial x} - \beta \nu' = 0 \quad (2.15)$$

the solutions are:

$$\psi(x, y, t) = \text{Re} \left[ \tilde{\psi} \sin(l y) e^{i(kx - \omega t)} \right] \quad (2.16)$$

as<sup>3</sup>

$$\zeta = \nabla^2 \psi = -(k^2 + l^2) \psi$$

with  $u = -il\psi$  and  $\nu = ik\psi$  the dispersion relation is

$$\omega = U k - \frac{k\beta}{(k^2 + l^2)}$$

---

<sup>3</sup>This will be understood after a reading of the Spectral Transformation method in sec.3.7.

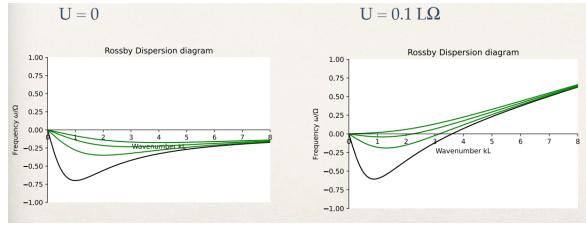


Figure 2.6: Rossby waves

## 2.4 Fundamental processes

Different climate models incorporate all the complex physical processes occurring in the climate system using a variety of methods. The dynamic involves interactions between motions, thermodynamics, and atmospheric water content.

### 2.4.1 Radiation

The radiation that heats and cools the climate system can be divided into two parts: solar radiation (shortwave) and terrestrial radiation (longwave). All gases emit and absorb radiation, and each chemical element or combination of elements has a distinct spectrum indicating at which electromagnetic wavelengths its emissions occur. In some cases, the emission may be confined to narrow portions of the electromagnetic spectrum rather than being a smoothly varying function of wavelength. This phenomenon of distinct, discrete emission spectra results from the emission of radiation as an electron moves from one orbit around an atomic nucleus to another orbit closer to the nucleus. If the atom is excited by absorbing energy, then the electron can go to a higher (outer) discrete orbit. As electrons return to the inner orbits, emission occurs in the ultraviolet part of the spectrum, producing the emission spectra. If electrons return to intermediate orbits instead of higher orbits, the emission is in the visible part of the spectrum; for the outer orbits, the emission is in the infrared. So far we have discussed only the emission of energy from a gas, not the absorption of energy. Suppose radiational energy enters an atomic or molecular gas and is absorbed. In that case, it can increase the atomic or molecular energy levels by the same amount of energy involved in the emission. This absorption can be just as discrete or selective as the emission. Thus, if radiation entering a gas cannot excite the atoms or molecules in any energy form, then the radiational energy will not be absorbed or emitted in the gas.

In the Earth's atmosphere, ozone, carbon dioxide, and water vapor are very important triatomic molecules that both emit and absorb radiation in certain parts of the electromagnetic spectrum that affect the climate system.

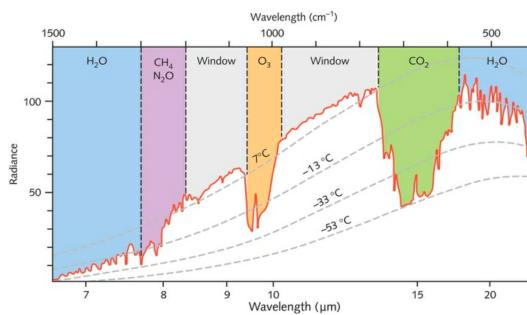


Figure 2.7: Earth's spectrum

As we can see in figure 2.7 ozone is a very strong absorber in the 9-10  $\mu\text{m}$  region, carbon dioxide has absorption maxima in the 2, 3, 4, and 13-17  $\mu\text{m}$  regions, and water vapor has several absorption regions in the 1-8  $\mu\text{m}$  range and at wavelengths greater than 13  $\mu\text{m}$ .

The radiational heating and cooling computation in climate models is usually done by calculating upward and downward fluxes through unit horizontal areas, considering the vertical distributions of temperature, water vapor, and other radiative absorbing gases such as carbon dioxide and ozone.

The blackbody radiation curve was determined theoretically by Max Planck in 1900. A result of Planck's equation is the earlier displacement law of Wien, that the wavelength of maximum intensity  $\frac{2897.8}{T}$  where  $T$  is the temperature of the blackbody. For the Earth, the approximate global mean surface temperature is 293 K, which yields a maximum intensity near 9.9  $\mu\text{m}$  whereas for the sun the surface temperature is approximately  $T = 6110\text{K}$  which yields a maximum intensity at 0.474  $\mu\text{m}$ . The overlap of the blackbody radiation curves for the Earth and for the sun's radiation reaching the Earth is not great, and consequently solar and terrestrial radiation can be separated based on wavelength.

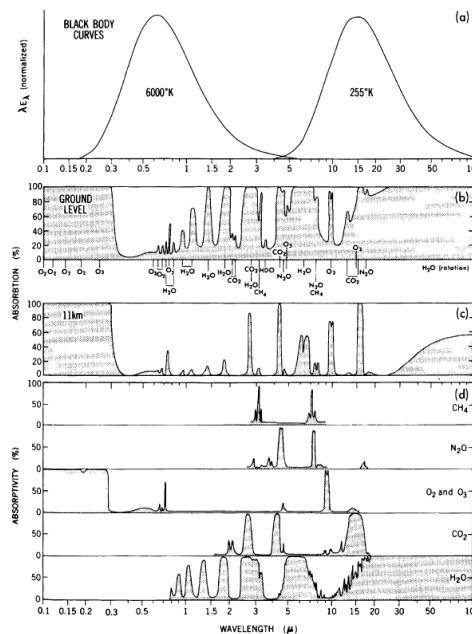


Figure 2.8: Black body curves for the solar radiation and the terrestrial radiation for the entire atmosphere and the first 11 km of the atmosphere

Much of the fundamental physics governing radiation transfer is embodied in the following two laws:

- Lambert's law, which provides a formulation for the decrease in intensity of radiation of a given wavelength as the radiation passes through a given amount of absorbing gas

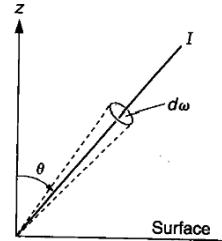
$$\frac{d}{dz}[I] = -Ik\rho$$

where  $k$  is the absorption coefficient (which tells us what kind of mass is preset),  $\rho$  is the density of the layer, and  $I$  is the radiance, defined as the energy per unit time per unit area per unit solid angle  $d\omega$

- Kirchhoff's law, states that there is a proportionality between radiative absorptivity and emissivity of a gas at the same temperature for any wavelength. A good absorber of radiation at some wavelength is also a good emitter of radiation at the same wavelength.

In particular, it is usually identified as *absorptivity* the ratio of the amount of radiative energy absorbed to the total incident radiation, and *emissivity* is the ratio of the emitted radiation to the maximum possible emitted radiation at the same temperature.

The 2.4.1 shows the relationship between intensity from some given direction, unit solid angle  $d\omega$ , and the surface, R0.38



such that integration over  $\theta$  of the hemisphere above the horizontal surface yields the flux,  $F$ , arriving from all angles:

$$F = \int (I \cos \theta) d\omega \quad (2.17)$$

Dividing 2.17 by  $I$  and forming integrals over optical depth on either side yield

$$\int \frac{dI}{I} = - \int k\rho dz$$

which becomes

$$\ln \left( \frac{I}{I_0} \right) = - \int k\rho dz$$

upon integrating the left-hand side. This can be written exponentially as

$$I = I_0 e^{-\chi}$$

$$\chi = \int k\rho dz$$

where  $\chi$  is the optical depth or optical path length (rate at which the radiance is degraded, if you move upward, at a certain point you get zero because you have traveled enough to absorb all the radiation) and

$$\tau = e^{-\chi}$$

is the frictional transmission, indicating the fraction of the original radiance that get transmitted to optical depth within attenuating gas. If  $\chi = 1$ , then  $\tau = 0.37$ , which means the initial intensity  $I$  is decreased by a factor of 0.63, for normal atmospheric conditions  $\chi$  is typically much less than 1.

For a blackbody, the emission is also described by Lambert's law with a different sign

$$dI = B(T)k\rho dz \quad (2.18)$$

Integrating 2.18 over a hemisphere, the Stefan-Boltzmann law is then obtained

$$\int_0^{2\pi} B(T) \cos \theta d\omega = \sigma T^4$$

or

$$\pi B(T) = \sigma T^4$$

The change of radiance at a point z is made up of two fluxes

$$\frac{1}{\rho} \frac{dI}{dz} = -k(I - B(T))$$

where  $kI$  is the absorption and  $kB(T)$  is the emission. So the fractional absorption in a layer  $(z, z_1)$  is

$$\tau(z, z_1) = \exp\left(- \int_{z_1}^z k\rho dz'\right)$$

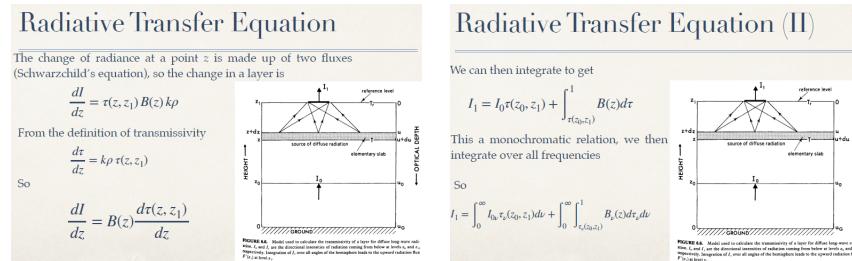
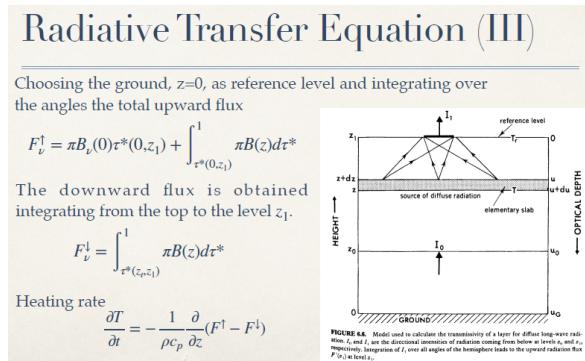


Figure 2.9: Non avevamo voglia di scriverli



**Solar radiation** The solar radiation is a function of the zenith angle

$$\cos Z = \sin \varphi \sin \delta + \cos \varphi \cos \delta \cos H$$

where  $\varphi$  is latitude,  $\delta$  is solar declination, which is the angular distance of the sun north of the equator and varies from about  $23.5^\circ$  on June 22 to  $-23.5^\circ$  on December 22 and  $H$  is the hour angle, which is the longitudinal distance from the point in question to the meridian of solar noon and therefore is 0 at any point experiencing solar noon. The solar flux,  $S$ , entering at the top of the atmosphere is a function of  $\cos Z$  and the distance from the sun to the Earth,  $d$ , such that

$$S = S_0 f(d) \cos Z$$

where  $S_0$  is the so-called solar constant (the solar energy flux received at the outer atmosphere on a surface normal to the solar beam; known now not to be constant) and the factor  $f(d)$  is 1.0344 in early January and 0.9674 in July for the present astronomical conditions. As shown in 2.10 the principal absorber in the atmosphere is the stratospheric ozone, which absorbs very effectively in the ultraviolet and in the visible. Water vapor is the primary absorber in the troposphere in the near-infrared, also with carbon dioxide are the main important absorbers for the longer wavelengths.

The largest contributors to the heating/cooling rates are water vapor, carbon dioxide and ozone as shown in Figure 2.11. The net heating/cooling is zero in the stratosphere because

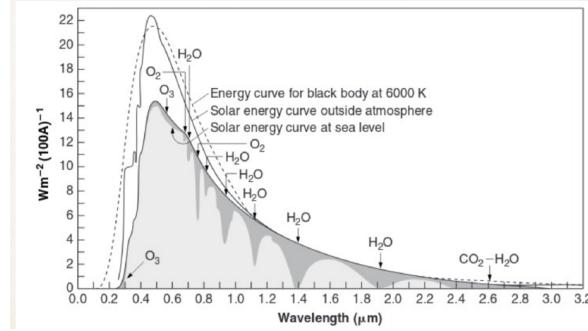


Figure 2.10: Spectral energy distribution as a function of wavelength at the top of the atmosphere and at sea level. The figure includes absorption by various atmospheric gases

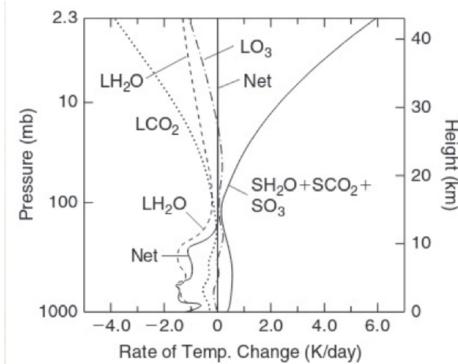


Figure 2.11: Heating rates in the atmosphere due to the absorption of solar radiation by atmospheric gases and due to the longwave or infrared radiation

in the one-dimensional model, it is in radiative equilibrium. The troposphere shows a net radiative cooling that is compensated by the vertical transfer of sensible and latent heat from below by moist adiabatic convection. As suggested in Figure 2.11 water vapor is the strongest contributor to the troposphere cooling. In the stratosphere cooling due to water vapor, carbon dioxide and ozone is generally compensated by heating due to absorption of solar radiation by zone.

#### 2.4.2 Moisture

In order to discuss precipitation and cloud physics used in climate models, let's introduce basic concepts. *Moisture* is the quantity of water vapor in the air and it can be specified in several ways, depending on which reference is used.

- *mixing ratio* ratio of the density of water vapor to the density of dry air  $q = \frac{\rho_w}{\rho}$
- *specific humidity* ratio concerning the density of moist air, the density of water vapor over the total density of the air:  $h = \frac{\rho_w}{\rho_w + \rho}$
- *relative humidity* ratio of the mixing ratio to the saturation mixing ratio  $r = \frac{q}{q_s}$

The same as the continuity mass treatment, the changes in the amount of water vapor must be balanced by the moisture sources and sinks

$$\frac{dq}{dt} = \frac{1}{\rho} M + E \quad (2.19)$$

where  $M$  is the time rate change of water vapor per unit volume due to condensation or freezing (sinks of moisture) and  $E$  is the time rate of change of water vapor content per unit

mass due to evaporation from the surface and sub-grid vertical and horizontal diffusion of moisture within the atmosphere (source of moisture). The first term on the right is a sink of moisture, the second is a source of moisture. Often 2.19 is written in flux form by combining with the continuity equation to obtain

$$\frac{\partial(\rho q)}{\partial t} + \nabla \cdot (\rho q \mathbf{V}) + \frac{\partial(\rho q \omega)}{\partial z} = M + \rho E \quad (2.20)$$

If 2.20 is integrated over the entire volume of the atmosphere, the second and the third terms on the left drop out, so that the sources and sinks of moisture must balance to have no secular change in atmospheric moisture over the globe. If the atmosphere is saturated with moisture, then sensible heat can be added to the atmospheric system from latent heat by the conversion of water vapor to liquid warmer or ice parcels.

The first law of thermodynamic must incorporate in the nonadiabatic term, this energy conversion process due to phase changes between liquid, solid, and gas. If the nonadiabatic process of conversion of water vapor to liquid water is the only energy source incorporated, the first law of thermodynamics becomes

$$c_p dT + \frac{1}{\rho} dp = -L dqs$$

The equation of state can be written like

$$e = \rho_w RT$$

where  $e$  is the partial pressure for water vapor. Substituting in the previous equation, the mixing ratio becomes now

$$q \approx 0.622 \frac{e}{p} \quad (2.21)$$

Inserting a total differential form of the hydrostatic equation, the first law of thermodynamics becomes

$$\frac{dT}{dz} = -\frac{g}{c_p} - \frac{L}{c_p} \frac{dqs}{dz}$$

taking the log and differentiating the previous equation 2.21 we will obtain

$$\begin{aligned} \log q_s &\approx \log 0.622 + \log e_s - \log p \\ \frac{1}{e_s} \frac{de_s}{dz} &= \frac{1}{q_s} \frac{dq_s}{dz} + \frac{1}{p} \frac{dp}{dz} \\ \frac{1}{e_s} \frac{de_s}{dT} &= \frac{1}{q_s} \frac{dq_s}{dz} - \frac{1}{p} g \rho \end{aligned}$$

The Clausius-Clapeyron equation relates the change in saturation vapor pressure to the latent heat involved in a phase change from water vapor to liquid water or from liquid water to ice. The most convenient form of this is

$$\frac{1}{e_s} \frac{de_s}{dT} = \frac{L}{RT^2}$$

So we get

$$\begin{aligned} \frac{dq_s}{dz} &= \frac{L q_s}{R W T_2} \frac{dT}{dz} + \frac{g q_s}{R T} \\ \Gamma_m &= \frac{dT}{dz} = -\frac{g}{c_p} \left( 1 + \frac{L q_s}{R T} \right) \left( 1 + \frac{L^2 q_s}{c_p R W T_2} \right)^{-1} \end{aligned} \quad (2.22)$$

In this way, we will obtain the moist adiabatic lapse rate  $\Gamma_m$ , which is less than the adiabatic lapse rate defined earlier:

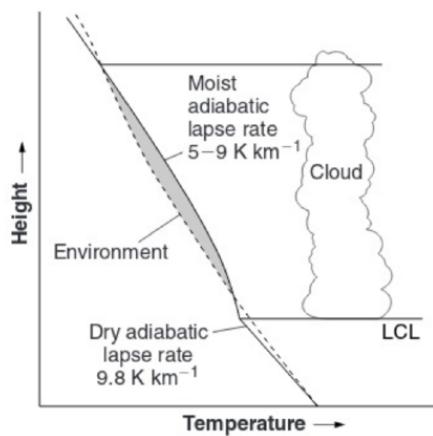
$$\Gamma = \frac{dT}{dz} = -\frac{k \rho g}{p} T$$

### 2.4.3 Clouds

They play a critical role in the radiation characteristics of the Earth's climate, but their radiative properties are not fully understood. In most climate models, the precipitation and convective aspects of the model are used to compute where clouds exist based on whether the atmosphere is near saturation and whether convection is occurring. You can use different parameters to explain that.

#### Cumulus formation

If a parcel of air near the Earth's surface (at an atmospheric pressure of approximately 1000 mb) starts to rise, say from strong heating, it will cool at approximately the dry adiabatic lapse rate of  $9.8 \text{ K km}^{-1}$ . When the parcel cools sufficiently the parcel is saturated, gaining a condition of instability because there is much more energy in the water vapor. The level at which this occurs is termed the lifting condensation level, LCL, and is typically at the cloud base as the picture below shows. If the parcel does not rise high enough to reach an



LCL, then a cloud should not form. From the LCL upward, the parcel will cool less rapidly as it rises since the latent heat of condensation or fusion releases heat. This lessened rate of temperature decrease with height is the moist adiabatic lapse rate.

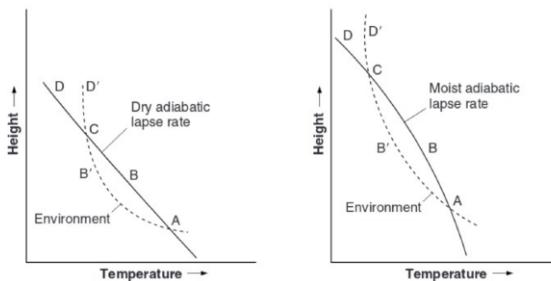


Figure 2.12: Diagram of stable and unstable conditions concerning dry and moist lapse rates.

As shown in figure 2.12 the dotted curve represents conditions for the environmental air, and the solid curves represent conditions for the path along which a parcel will ascend from point A to B, depending on whether the dry or the moist adiabatic lapse rate is appropriate. There is a point B' highlighted on the curve for the environmental air that is colder in both cases than point B for the parcel at the same height. Since the air at B' is colder than that at B, the parcel will have positive buoyancy. Thus in principle, it will continue ascending. If on the other hand, a parcel is at C and ascends to point D, it will be colder than the surrounding

environmental air at D'. Thus the parcel will have negative buoyancy, which will cause it to sink to its original position.

#### 2.4.4 Surface Processes

Boundary fluxes for momentum and heat, we can physically represent as

$$\tau_x = \rho C_D |\mathbf{v} - \mathbf{v}_s| (u_s - u)$$

$$\tau_y = \rho C_D |\mathbf{v} - \mathbf{v}_s| (v_s - v)$$

$$H = \rho c_p C_H |\mathbf{v} - \mathbf{v}_s| (\theta_s - \theta)$$

$$LE = \rho L_C E |\mathbf{v} - \mathbf{v}_s| (q_s - q)$$

Where over land  $\mathbf{v}$  is zero.

Above the surface boundary layer adjacent to Earth's surface exists the planetary boundary layer, where the wind turns with the height. Within this region, the balance of forces can be approximated by Coriolis, pressure gradient, and frictional forces through the following simplification

$$-f_\nu = -\frac{1}{\rho a \cos \varphi} \frac{\partial p}{\partial \lambda} + F_\lambda$$

$$f_u = -\frac{1}{\rho a} \frac{\partial p}{\partial \varphi} + F_\varphi$$

The frictional term can be expressed as the vertical gradient of stress, so that

$$F_\lambda = \frac{1}{\rho} \frac{\partial \tau_\lambda}{\partial z}$$

and

$$F_\varphi = \frac{1}{\rho} \frac{\partial \tau_\varphi}{\partial z}$$

and the stress can in turn be related to the vertical gradient of wind shear

$$\tau_\lambda = \rho K_m \frac{\partial u}{\partial z}$$

and

$$\tau_\varphi = \rho K_m \frac{\partial v}{\partial z}$$

where  $K_m$  is the vertical eddy or transfer coefficient for momentum. Note that if the vertical wind shear is large, as one would expect near the Earth's surface since the wind must approach zero at the surface, then the momentum transfer is also large. On the molecular scale,  $K_m$ , takes on values appropriate for those space scales; however, in the atmosphere and ocean models the space scales are much larger. Many of the eddies have dimensions of the order of the spatial grid structure of the models. If the layer is unstable there is strong vertical mixing and  $K_m$  is large; if the layer is stable there is no strong coupling and  $K_m$  is small. A useful measure of this stability is the Richardson number  $Ri$ , which is defined as

$$R_i = \frac{g}{\theta} \frac{\frac{\partial \theta}{\partial z}}{\left( \frac{\partial V}{\partial z} \right)^2}$$

where  $V = |\mathbf{V}|$  and  $\frac{\partial V}{\partial z}$  is the vertical wind shear or vertical gradient of wind velocity.

$$\begin{aligned}\frac{\partial \zeta}{\partial t} &= - \nabla \cdot (\zeta + f) \vec{V} \\ \frac{\partial \zeta}{\partial t} &= - (\zeta + f) \nabla \cdot \vec{V} - \vec{V} \cdot \nabla (\zeta + f) \\ \frac{\partial \zeta}{\partial t} &= - (\zeta + f) D - \vec{V} \cdot \nabla (\zeta + f)\end{aligned}$$

Figure 2.13: Schematic processes and features relevant to surface hydrology and evapotranspiration

#### 2.4.5 Hydrology

As shown in 2.13 lots of processes are involved in surface hydrology. These include radiation, sensible heat, evaporation, transpiration, momentum, and precipitation, but also H<sub>2</sub>O and CO<sub>2</sub> fluxes. The surface moisture can contribute to surface flow in streams percolate to a deeper layer or even become groundwater, depending on the porous properties of the ground. In the root zones of plants, some of the surface moisture can be taken up by the plant roots and then relayed to the plant leaves and transpired.

Within a land area as large as a horizontal grid cell of a global model, there are generally multiple surface types, including lakes, wetlands, different vegetation types, and different soil types. Another aspect of land surface processes being improved in many climate models is the simulation of snow coverage. The snow modeling community has developed a variety of models that can be used for climate and hydrology. These models range from the simple to the complex, and several of them are specifically designed for use in general circulation models. The more complex snow models take into account melting processes, grain size, shape, liquid water content, and percolation processes.

# Chapter 3

## Numerical methods used in atmospheric models

Last updated: 2024-12-08  
Source file: [chapter-numerical-methods.tex](#)

*Major content of this chapter is taken from prof. Navarra's teaching resources, available in his website [https://wanderer.cmcc.it/index\\_GARP.html](https://wanderer.cmcc.it/index_GARP.html)*

### 3.1 Historical introduction

In this chapter, following a short historical introduction on the development and use of numerical methods in atmospheric models, methods available for numerical solution of the differential equations governing the atmosphere will be briefly reviewed. Then, basic elements of the finite difference method for solving these equations will be introduced. Finally, the concept of stability of finite difference equations, and methods for testing the stability of such equations, will be discussed at some length.

It is considered that Wilhelm Bjerknes (1904) was the first to point out that the future state of the atmosphere can in principle be obtained by an integration of differential equations which govern the behavior of the atmosphere, using as initial values fields describing an observed state of the atmosphere. Such an integration performed using numerical methods is called *numerical weather prediction*. When, however, a numerical integration is performed starting from fictitious initial fields, it is called *numerical simulation*.

A first practical attempt at a numerical weather prediction was made by Richardson. After very tedious and time-consuming computations, carried out mostly during the First World War, Richardson obtained a totally unacceptable result. Despite this, he described his method and results in a book (Richardson, 1922), and this is today one of the most famous in meteorology.

The wrong result obtained by Richardson, and his estimate that 64,000 men are necessary to advance the calculations as fast as the weather itself is advancing, left some doubt as to whether the method would be of practical use. A number of developments that followed, however, improved the situation. Courant, Friedrichs and Lewy (1928) found that space and time increments in integrations of this type have to meet a certain stability criterion. Mainly due to the work of Rossby in the late 1930's, it became understood that even a rather simple equation, that describing the conservation of absolute vorticity following the motion of air particles, suffices for an approximate description of large-scale motions of the atmosphere. Finally, in 1945, the first electronic computer ENIAC (Electronic Numerical Integrator and Computer) was constructed. The absolute vorticity conservation equation, and this first electronic computer, were used by Charney, Fjortoft and von Neumann in the late 1940's for the first successful numerical forecast (Charney et al., 1950).

Much faster computers, and improved understanding of computational problems, now also enable long-term integrations of the basic primitive equations. It is generally considered that integration of the primitive equations enables easier incorporation of various physical processes than the integration of modified equations, that is, integration of the divergence and vorticity equations. Thus, it is mostly the primitive equations that are used today for practical numerical forecasting by meteorological services. Charts obtained by numerical forecasting are used by synopticians in these services as the principal basis for decisions on forecasts issued for public use.

A number of research groups have been actively engaged for more than a decade in development of models for the numerical simulation of the general circulation of the atmosphere. In such simulations starting from a fictitious initial state, e.g. an isothermal and motionless atmosphere, is often considered to be an advantage for the experiments. It enables a test of the ability of the computational and physical schemes of the model to simulate an atmosphere with statistical properties similar to those of the real atmosphere, with no, or not much, prior information on these properties.

Numerical models are also very frequently developed for studies of some smaller-scale atmospheric phenomena. Foremost among these are studies of the cumulus convection problem, and simulation of processes within the planetary boundary layer. In this text, however, we shall primarily have in mind the application of numerical methods to prediction and simulation of large-scale atmospheric motions.

### 3.2 Methods for the numerical solution of the equations of motion

Numerical solution of the equations of motion today in most cases is performed using the *grid point method*. In this method a set of points is introduced in the region of interest and dependent variables are initially defined and subsequently computed at these points. This set of points is called the grid. The words mesh or lattice are also used. It is necessary to have the grid points at fixed locations in the horizontal. This means that, according to the Eulerian system of equations, space and time coordinates are chosen as independent variables.

A number of attempts have been made to develop atmospheric models using an approach which is at least partly Lagrangian<sup>1</sup>. Serious difficulties are encountered when a straightforward numerical integration of the Lagrangian system of equations is undertaken. However, it is possible to construct methods with some Lagrangian properties; for example, to have some or all of the computation points moving with the fluid. In hydrodynamics a number of such methods have proved to be very useful, especially for some problems which are not amenable to treatment by a strictly Eulerian technique (e.g. Harlow and Amsden, 1971). However, in meteorology the performance of Lagrangian or semi-Lagrangian models that have so far been developed has not been quite satisfactory. A discussion of one way of constructing a Lagrangian model, and a review of earlier attempts, can be found in a paper by Mesinger (1971).

Another possible approach is to express the spatial dependence of the variables in terms of a series of orthogonal functions, and then substitute this into the governing equations. In this way the equations reduce to a set of ordinary differential equations, so that the coefficients of the series can be computed as functions of time. This is the *spectral method* of solving the governing equations. Until relatively recently it was considered that in efficiency the spectral

---

<sup>1</sup>The Eulerian approach focuses on observing how a field (like fluid flow) changes at fixed points in space over time, much like watching traffic from a fixed camera. In contrast, the Lagrangian approach tracks individual elements or particles as they move through space and time, similar to following a specific car on its journey. Essentially, Eulerian examines the "big picture" from a stationary perspective, while Lagrangian follows individual components in motion.

method could not be competitive with the grid point method. But the use of the fast Fourier transform has completely changed the situation and investigation of spectral methods is now the subject of intensive research.

In the following we shall consider the technique of using the grid point method, and the problems associated with it, using grid of computation points fixed in space. This is the most direct way of solving the equations of motion numerically. Furthermore, knowledge of this method is necessary for the investigation and understanding of the relative merits of other alternatives mentioned in this section.

### 3.3 Grid point method

With the grid point method, the most common way of solving the governing equations is to find approximate expressions for derivatives appearing in the equations. These approximate expressions are defined using only values of the dependent variables at the grid points, and at discrete time intervals. Thus, they are formed using differences of dependent variables over finite space and time intervals; for that reason this approach is called the *finite difference method*. The approximations for derivatives are then used to construct a system of algebraic equations that approximates the governing partial differential equations. This algebraic system is considered valid at each of the interior grid points of the computation region. For the initial time and at space boundary points, additional constraints or equations are defined that approximate the initial and boundary conditions as required by the physics of the problem. The set of algebraic equations obtained in this way is then solved, usually using an electronic computer, by a suitable step-wise procedure.

We shall now consider some basic elements of the finite difference method. For simplicity, we start by considering a function of one independent variable

$$u = u(x)$$

The function  $u$  is a solution to a differential equation that we are interested in. We want to find an approximation to this solution in a bounded region  $R$  of the independent variable, having a length  $L$ . The simplest way of introducing a set of grid points is to require that they divide the region  $R$  into an integer number of intervals of equal length  $\Delta x$ . This length  $\Delta x$  is called the *grid interval*, or *grid length*. Let us denote the number of grid intervals by  $J$ . It is convenient to locate the origin of the  $x$  axis at the left-hand end of the region  $R$ . Thus, we are looking for approximations to  $u(x)$  at discrete points  $x = j\Delta x$ , where  $j$  takes integer values  $0, 1, 2, \dots, J$ . These approximate values we shall denote by

$$u_j = u_j(j\Delta x)$$

Thus, we are interested in finding  $J + 1$  values  $u_j$

Knowledge of a discrete set of values  $u_j$ , even if the approximations were perfect, offers, obviously, less information than knowledge of the function  $u(x)$ . Let us briefly consider the situation in that respect. We shall very often find it convenient to think of the function  $u(x)$  as being formed by a sum of its Fourier components, that is

$$u(x) = \frac{a_0}{2} + \sum_{n \geq 1} [a_n \cos(2\pi n \frac{x}{L}) + b_n \sin(2\pi n \frac{x}{L})]$$

Now, the available  $J + 1$  values  $u_j$  do not enable the computation of all of the coefficients  $a_n b_n$ ; rather, they can be used to compute only  $J + 1$  different coefficients. A natural choice is to assume that the  $J + 1$  values  $u_j$  define the near value  $a_0$  and as many as possible of the coefficients of the Fourier components at the long wave length end of the series, that is,

coefficients for  $n = 1, 2, 3, \dots, J/2$ . Of these components, the one with the shortest wavelength will have  $n = J/2$ , with the wave length

$$\frac{L}{n} = \frac{2L}{j} = \frac{2L}{\frac{L}{\Delta x}} = 2\Delta x$$

Having made that choice, we can say that with values  $u_j$  at discrete points  $x = j\Delta x$  it is not possible to resolve waves with wave length shorter than  $2\Delta x$ . Now let us consider the differences between values  $u$  that will be used to construct approximations to derivatives. These differences are called *finite differences*. They can be calculated over one or more of the intervals  $\Delta x$ . Depending on the relation of the points from which the values are taken to the point where the derivative is required, they can be *centered* or *uncentered*. An un-centered difference is, for example, the *forward* difference

$$\Delta u_j = u_{j+1} - u_j \quad (3.1)$$

more often centered (or central) differences are used, such as

$$\Delta u_{j+1/2} = u_{j+1} - u_j \quad (3.2)$$

In a centered difference the difference is between values symmetrical about the point where the difference is being calculated.

One way to construct an approximation to a differential equation is to simply replace the derivatives by appropriate *finite difference quotients*. For example, for the first derivative one can use the approximation (in a forward difference scheme):

$$\left( \frac{du}{dx} \right)_j \rightarrow \frac{u_{j+1} - u_j}{\Delta x} \quad (3.3)$$

The finite difference quotient here is, of course, only one of many possible approximations to the first derivative at point  $j$ .

If a finite difference quotient, or a more complex expression, is to be used as an approximation to a derivative, it is required, above all, that this approximation be consistent. This means that the approximation should approach the derivative when the grid interval approaches zero. The quotient 3.3, obviously, has that property.

Important information is obtained when the true solution  $u(j\Delta x)$  is substituted into an approximation to the derivative in place of the grid point values  $u_j$ , and  $u(j\Delta x)$  is expanded in a Taylor series about the central point. For the quotient 3.3 this procedure gives

$$\frac{u_{j+1} - u_j}{\Delta x} \rightarrow \left( \frac{du}{dx} \right)_j + \frac{1}{2} \left( \frac{d^2 u}{dx^2} \right)_j \Delta x + \frac{1}{6} \left( \frac{d^3 u}{dx^3} \right)_j (\Delta x)^2 + \dots \quad (3.4)$$

The difference between this expression and the derivative

$$\left( \frac{du}{dx} \right)_j$$

being approximated. In this case:

$$\epsilon = \frac{1}{2} \left( \frac{d^2 u}{dx^2} \right)_j \Delta x + \frac{1}{6} \left( \frac{d^3 u}{dx^3} \right)_j (\Delta x)^2 \quad (3.5)$$

is called the *truncation error* of the approximation to the derivative. These are terms that were "truncated off" to form the approximation. The truncation error gives a measure of how accurately the difference quotient approximates the derivative for small values of  $\Delta x$ . The usual measure of this is the order of accuracy of an approximation. This is the lowest power

of  $\Delta x$  that appears in the truncation error. Thus, approximation 3.3 is of the first order of accuracy. We can write

$$\epsilon = o(\Delta x) \quad (3.6)$$

For an approximation to the derivative to be consistent it must, obviously, be at least first order accurate.

### 3.3.1 Finite difference schemes

The algebraic equation obtained when derivatives in a differential equation are replaced by appropriate finite difference approximations is called a finite difference approximation to that differential equation, or a finite difference scheme. In this section we shall introduce the concepts of consistency, truncation error, and accuracy, for a finite difference scheme.

As an example, we shall use the linear advection equation

$$\begin{aligned} \frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} &= 0 \\ u &= u(x, t) \end{aligned} \quad (3.7)$$

$c$  is a positive constant. It describes advection of the variable  $u$  at a constant velocity  $c$  in the direction of the  $x$  axis. The solution to this simple equation can, of course, also be obtained by an analytic method. It will be useful to obtain the analytic solution first, in order to investigate properties of numerical solutions by comparing them against known properties of the true solution.

It is convenient to this end to change from variables  $x, t$  to variables  $\xi$  and  $t$  with the substitution  $\xi = x - ct$ .

Using the notation

$$u^{(x,t)} = U(\xi, t)$$

we obtain:

$$\frac{\partial u}{\partial t} = \frac{\partial U}{\partial \xi} \frac{\partial \xi}{\partial t} + \frac{\partial U}{\partial t} \frac{\partial t}{\partial t} = -c \frac{\partial U}{\partial \xi} + \frac{\partial U}{\partial t} \quad (3.8)$$

$$\frac{\partial u}{\partial x} = \frac{\partial U}{\partial \xi} \frac{\partial \xi}{\partial x} + \frac{\partial U}{\partial t} \frac{\partial t}{\partial x} = \frac{\partial U}{\partial \xi} \quad (3.9)$$

that, substituted into 3.7, gives

$$\frac{\partial}{\partial t} U(\xi, t) = 0$$

Thus, it is seen that  $U$  cannot be a function of  $t$ , but can be an arbitrary function of  $\xi$ . A solution of 3.7 is, therefore,

$$u = f(x - ct) \quad (3.10)$$

where  $f$  is an arbitrary function. This, we see, is the general solution of the advection equation 3.7, since it can satisfy an arbitrary initial condition

$$u(x, 0) = F(x) \quad (3.11)$$

Thus,

$$u = F(x - ct) \quad (3.12)$$

is the solution of 3.7 satisfying the initial conditions. For a physical interpretation, it is often convenient to consider the solution in the  $x, t$  plane. In the present case, we see that the solution takes constant values along the straight lines

$$x - ct = \text{const}$$

These lines are the *characteristics* of the advection equation. We can say that the solution propagates along the characteristics.

Let us now construct a scheme for finding an approximate solution to 3.7 using the grid point method.

We are now looking only for an approximate solution at the discrete points in the  $(x, t)$  plane formed by the grid shown in the figure below. The approximate solution at a point  $(j\Delta x, n\Delta t)$  is denoted by  $u_j^n$ .

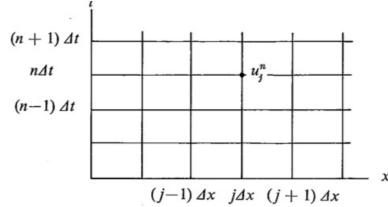


Figure 3.1: A finite difference grid for finding an approximate solution of 3.7

The behaviour of the true solution, which propagates along characteristics in the  $(x, t)$  plane, suggests constructing the approximate equation by replacing the time derivative by a forward difference quotient, and the space derivative by a backward difference quotient. In this way we obtain the scheme

$$\frac{u_j^{n+1} - u_j^n}{\Delta t} + c \frac{u_j^n - u_{j-1}^n}{\Delta x} = 0 \quad (3.13)$$

This scheme could be called a forward and upstream scheme, the latter word indicating the position of the point  $j - 1$  relative to the advection velocity. It is, of course, only one of many possible consistent finite difference schemes for the differential equation. There are many schemes which approach the differential equation when the increments  $\Delta x, \Delta t$  approach zero.

Since for small values of  $\Delta x, \Delta t$  a finite difference equation approximates the corresponding differential equation, we can expect that its solution will be an approximation to the solution of that equation. We shall call solutions given by finite difference schemes numerical solutions. There are, of course, both approximate and numerical solutions obtained by other methods which will not be considered in this publication. It is most convenient to study the properties of numerical solutions when they can be compared with known solutions of the original differential equation, which we shall refer to as true solutions. The difference between the numerical and the true solution

$$u_j^n - u(j\Delta x, n\Delta t) \quad (3.14)$$

is the error of the numerical solution. For obvious reasons, we cannot often expect to know the error of the numerical solution. However, we can always find a measure of the accuracy of the scheme by substituting the true solution  $u(j\Delta x, n\Delta t)$  of the equation, into the numerical scheme. Since the true solution will not satisfy the numerical equations exactly, we will have to add an additional term to keep the equation valid. Let us denote this term by  $\epsilon$ . For example, in the case of scheme 3.13 this procedure gives

$$\frac{u(j\Delta x, (n+1)\Delta t) - u(j\Delta x, n\Delta t)}{\Delta t} + c \frac{u(j\Delta x, n\Delta t) - u((j-1)\Delta x, \Delta t)}{\Delta x} = \epsilon \quad (3.15)$$

The term  $\epsilon$  we shall call the truncation error of the finite difference scheme. It shows how closely the true solution satisfies the equation of the scheme, and, thus, gives a measure of the accuracy of the scheme.

We can obtain a more useful form for the expression for the truncation error by performing a Taylor series expansion of the true solution about the central space and time point. Using the original differential equation to eliminate the leading term we obtain the truncation error 3.15:

$$\epsilon = \frac{1}{2} \frac{\partial^2 u}{\partial t^2} \Delta t + \frac{1}{6} \frac{\partial^3 u}{\partial t^3} (\Delta t)^2 + \dots - c \left( \frac{1}{2} \frac{\partial^2 u}{\partial x^2} \Delta x - \frac{1}{6} \frac{\partial^3 u}{\partial t^3} (\Delta x)^2 + \dots \right) \quad (3.16)$$

As before, these are the terms that were “truncated off” to make the differential equation reduce to our finite difference scheme.

In the same way as for an approximation to the derivative, the order of accuracy of a finite difference scheme is the lowest power of  $\Delta x$  and  $\Delta t$  that appears in the truncation error. Thus, scheme 3.13 is first order accurate. We can write

$$\epsilon = o(\Delta t) + o(\Delta x) = o(\Delta x, \Delta t)$$

It is useful to make a distinction between orders of accuracy in space and in time, especially when the lowest powers of  $\Delta x$  and  $\Delta t$  are not the same. As before, a necessary condition for consistency of a scheme is that it be at least of the first order of accuracy.

### 3.3.2 Convergence

The truncation error of a consistent scheme can be made arbitrarily small by a sufficient reduction of the increments  $\Delta x$  and  $\Delta t$ . Unfortunately, we cannot be sure that this will also result in a reduction of the error of the numerical solution. For that reason, we return to consideration of the error

$$u_j^n - u(j\Delta x, n\Delta t)$$

Following Richtmyer and Morton (1967) we ask two questions:

- What is the behaviour of the error  $u_j^n - u(j\Delta x, n\Delta t)$  when, for a fixed total time  $n\Delta t$  the increments  $\Delta x, \Delta t$  approach zero?
- What is the behaviour of the error  $u_j^n - u(j\Delta x, n\Delta t)$  when, for fixed values of  $\Delta x, \Delta t$ , the number of time steps  $n$  increases?

The answer to the first of these questions depends on the convergence of the numerical solution: if the error approaches zero as the grid is refined (as  $\Delta x \Delta t \rightarrow 0$ ) the solution is called convergent. If a scheme gives a convergent solution for any initial conditions, then the scheme also is called convergent.

Consistency of a scheme does not guarantee convergence. We shall illustrate this by a simple example. We still consider the scheme 3.13; its truncation error 3.16 approaches zero as the grid is refined, and, therefore, this is a consistent scheme. But consider the numerical solution, when the grid lines and characteristics are as shown in the figure below. The characteristic passing through the grid point taken as the origin in this example passes through another grid point, A, denoted by a square. Thus, the true solution at A, is equal to the initial value at the origin. However the numerical solution given by 3.13 at A is computed using the values at points denoted by circles. The shaded domain, including all of these points, is called *the domain of dependence* of the numerical scheme. The grid point at the origin is outside that domain, and, thus, cannot affect the numerical solution at A. Therefore, the error can be arbitrarily great. If the space and time steps were reduced by the same relative amount, say to one half of their values in the figure, the domain of dependence would still remain the same, and this situation would not change. Thus, as long as the ratio of the steps  $\Delta x$  and  $\Delta t$  remains the same, refinement of the grid cannot bring about a reduction in the error of the numerical solution.

$$x - ct = \text{const}$$

A necessary condition for convergence of a scheme is, obviously, that the characteristic defining the true solution at a grid point is inside the domain of dependence of the numerical solution at that point. In our example, this will happen when the slope of the characteristics is greater than the slope of the dashed line bounding the domain of dependence, that is, when

$$c\Delta t \leq \Delta x$$

thus, this is a necessary condition for convergence of 3.13

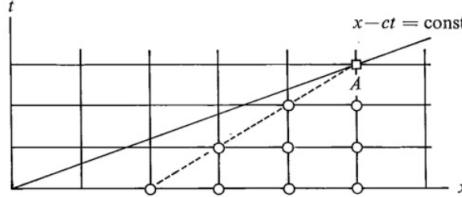


Figure 3.2: A possible relative position of a characteristic and of a domain of dependence

### 3.3.3 Stability

The answer to the second question raised at the beginning of the section Convergence "What is the behaviour of the error  $u_j^n - u(j\Delta x, n\Delta t)$  when, for fixed values of  $\Delta x, \Delta t$ , the number of time steps  $n$  increases?" depends on the stability of the numerical solution. A rigorous definition of stability employs the concepts of functional analysis, and refers to the boundedness of the numerical solution only (e.g. Richtmyer and Morton, 1967). The difficulties in defining stability are caused by the fact that the true solution, in general, does not have to be bounded. However, when we know that the true solution is bounded, as in the equations we are interested in here, we can use a definition referring to the boundedness of the error  $u_j^n - u(j\Delta x, n\Delta t)$ . We said that a solution  $u_j^n$  is stable if this error remains bounded as  $n$  increases, for fixed values of  $\Delta x, \Delta t$ . As before, we say that **a finite difference scheme is stable if it gives a stable solution for any initial condition**.

Stability of a scheme is a property of great practical significance. There are consistent schemes, of a high order of accuracy, that still give solutions diverging unacceptably fast from the true solution. Thus, conditions for stability, if any, should be known. There are three methods that can be used to investigate the stability of a scheme, and we shall give an example of each of these methods. We shall do this by considering again the forward and upstream scheme 3.13.

*Direct method* tests mathematically the stability of a scheme. Since we know that the true solution is bounded, it suffices to test the boundedness of the numerical solution. The scheme 3.13 can be written as

$$u_j^{n+1} = (1 + \mu)u_j^n + \mu u_{j-1}^n \quad (3.17)$$

where

$$\mu = c \frac{\Delta t}{\Delta x}$$

if  $1 - \mu \geq 0$ , which happens to be also the necessary condition for convergence, we will have:

$$|u_j^{n+1}| \leq (1 - \mu)|u_j^n| + \mu|u_{j-1}^n| \quad (3.18)$$

We can apply this at the point where at time level  $n+1$ ,  $|u_j^{n+1}|$  is a maximum,  $\text{Max}_{(j)}|u_j^{n+1}|$ . The right side of 3.18 can only be increased by replacing  $|u_j^n|$  and  $|u_{j-1}^n|$  by the maximum value at level  $n$ ,  $\text{Max}_{(j)}|u_j^n|$ . The two terms on the right side can be added, and we obtain

$$\text{Max}_{(j)}|u_j^{n+1}| \leq \text{Max}_{(j)}|u_j^n|$$

This proves the boundedness of the numerical solution. Hence,  $1 - \mu \geq 0$  is seen to be a sufficient condition for stability of 3.17.

This direct testing of the stability is simple. Unfortunately, as might be anticipated from the argument, it is successful only for a rather limited number of schemes.

*Energy method.* This method is of a much wider applicability, and can be used even for nonlinear equations. If we know that the true solution is bounded, we test whether

$$\sum_j (u_j^n)^2$$

is also bounded. If it is, then every value  $u_j^n$  must be bounded, and the stability of the scheme has been proved. The method is called the energy method since in physical applications  $u^2$  is often proportional to some form of energy. Of course, there are examples when this is not so.

*Von Neumann's*, or the Fourier series method is the most frequently used method of linear

Squaring 2.6.1 and summing over  $j$  we obtain

$$\sum_j (u_j^{n+1})^2 = \sum_j [(1 - \mu)^2 (u_j^n)^2 + 2\mu(1 - \mu) u_j^n u_{j-1}^n + \mu^2 (u_{j-1}^n)^2]$$

We shall assume a cyclic boundary condition, for example  $u_{-1} \equiv u_j$ , then

$$\sum_j (u_{j-1}^n)^2 = \sum_j (u_j^n)^2$$

Now, using Schwarz's inequality

$$\sum ab \leq \sqrt{\sum a^2} \sqrt{\sum b^2}$$

and 2.6.4, we can write

$$\sum_j u_j^n u_{j-1}^n \leq \sqrt{\sum_j (u_j^n)^2} \sqrt{\sum_j (u_{j-1}^n)^2} = \sum_j (u_j^n)^2$$

Using 2.6.4 and 2.6.5 we see that, if  $1 - \mu \geq 0$ , 2.6.3 gives the inequality

$$\sum_j (u_j^{n+1})^2 \leq [(1 + \mu)^2 + 2\mu(1 - \mu) + \mu^2] \sum_j (u_j^n)^2$$

or

$$\sum_j (u_j^{n+1})^2 \leq \sum_j (u_j^n)^2$$

equations that can be expressed in form of a Fourier series, where each harmonic component is also a solution. Thus, we can test the stability of a single harmonic solution ; stability of all admissible harmonics will then be a necessary condition for stability of the scheme. A harmonic solution of the finite difference advection equation.

The Von Neumann condition permits exponential growth of the solution, but no faster. This is useful for analyzing cases where the true solution grows exponentially. However, when the true solution does not grow, as in example 2.6.7, it is common to replace the von Neumann condition with a sufficient condition, much less generous than that required by the original definition of stability :  $|\lambda| \leq 1$ .

The graphs of the  $|\lambda|^2$  curves are shown in the picture above. In general, as the wave length  $L$  increases, that is, as  $k$  approaches to zero, the amplification factor approaches unity for any value of the parameter  $\mu$ . The figure shows that within the stable region the scheme is damping for all values  $\mu \leq 1$ . The damping increases as the wave length decreases. Since the true solution has a constant amplitude, this damping reveals an error due to finite differencing. We can see that this error increases as the wave length decreases. At the shortest resolvable

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0$$

in the form of a single harmonic

$$u(x, t) = \operatorname{Re} [U(t)e^{ikx}] \quad (2.6.6)$$

Here  $U(t)$  is the wave amplitude, and  $k$  the wave number. Substituting this into the preceding equation we obtain

$$\frac{dU}{dt} + ikU = 0$$

Thus, the problem of solving a partial differential equation has been reduced to that of solving this ordinary differential equation. Its solution is

$$U(t) = U(0)e^{-ikt}$$

where  $U(0)$  is the initial value of the amplitude. Hence, the desired harmonic solution is

$$u(x, t) = \operatorname{Re} [U(0)e^{ik(x-ct)}] \quad (2.6.7)$$

Each wave component is, thus, advected at a constant velocity  $c$  along the  $x$  axis with no change in amplitude.

$$u_j^n = \operatorname{Re} [U^{(n)}e^{ikj\Delta x}] \quad (2.6.8)$$

Here  $U^n$  is the amplitude at time level  $n$ . This substitution shows that 2.6.8 is a solution provided that

$$U^{n+1} = (1 - \mu)U^{(n)} + \mu U^{(n)}e^{-ik\Delta x} \quad (2.6.9)$$

An equation of this kind enables analysis of the behavior of the amplitude  $U^n$  as  $n$  increases.

To this end we define an *amplification factor*  $|\lambda|$  by

$$U^{n+1} \equiv \lambda U^n \quad (2.6.10)$$

This gives

$$|U^{n+1}| = |\lambda| |U^{(n)}|.$$

For each harmonic solution 2.6.8 to be stable it is required that

$$|U^n| = |\lambda|^n |U^{(0)}| \leq B$$

where  $B$  is a finite number. This gives

$$n \log |\lambda| \leq \log \frac{B}{U^0} \equiv B'$$

where  $B'$  is a new constant. Since  $n = \frac{t}{\Delta t}$ , the necessary condition for stability becomes

$$\log |\lambda| \leq \frac{B'}{t} \Delta t \quad (2.6.11)$$

where  $B'$  is a new constant. Since  $n = \frac{t}{\Delta t}$ , the necessary condition for stability becomes

$$\log |\lambda| \leq \frac{B'}{t} \Delta t \quad (2.6.11)$$

Now, suppose that we require boundedness of the solution for a finite time  $t$ . Condition 2.6.11 can then be written as

$$\log |\lambda| \leq 0(\Delta t)$$

If we now define

$$|\lambda| \equiv 1 + \delta$$

we see, in view of the power series expansion of  $\log(1 + \delta)$ , that the stability condition obtained is equivalent to

$$\delta \leq 0(\Delta t)$$

or

$$|\lambda| \leq 1 + O(\Delta t) \quad (2.6.12)$$

This is the *von Neumann necessary condition for stability*.

Returning to our example, substitution of 2.6.10 into 2.6.9 gives

$$\lambda = 1 - \mu + \mu e^{-ik\Delta x} \quad (2.6.14)$$

From this we obtain

$$\lambda^2 = 1 - 2\mu(1 - \mu)(1 - \cos k\Delta x) \quad (2.6.15)$$

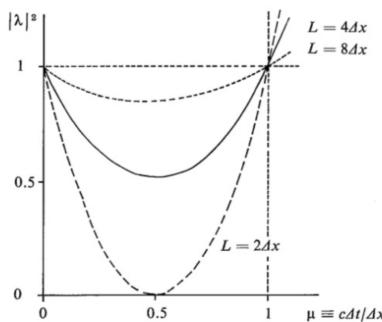
and, therefore,  $1 - \mu \geq 0$ .

An equation such as 2.6.15 gives further information about the behavior of the numerical solution. This can be obtained by studying the variation of  $|\lambda|$  with  $\mu$  for various fixed values of  $k\Delta x$ . To this end we plot the  $|\lambda|^2$  curves; 2.6.15 shows that in the present case all of these curves are parabolas. Furthermore, recall that the minimum resolvable wave length is  $2\Delta x$ . Thus, the maximum value that wave number  $k$  can take is  $\frac{\pi}{\Delta x}$ . We thus plot the  $|\lambda|^2$  curves for this maximum value  $k = \frac{\pi}{\Delta x}$  (or wave length  $L = 2\Delta x$ , and for half this value,  $k = \frac{\pi}{2\Delta x(L-4\Delta x)}$ , and a quarter of this value,  $k = \frac{\pi}{4\Delta x(L-8\Delta x)}$ ).

The first derivative

$$\frac{d|\lambda|^2}{d\mu} = -2(1 - 2\mu)(1 - \cos k\Delta x)$$

shows that all the  $|\lambda|^2$  curves have minima at  $\mu = \frac{1}{2}$ . This information, in addition to calculation of the ordinates of 2.6.15 at  $\mu = \frac{0.1}{2}$  and 1, suffices for sketching



2.6.1 Sketches of  $|\lambda|^2$  curves, for the scheme 2.6.1, and for various values of  $L$ .

wave length  $L = 2\Delta x$ , the error may be very great unless  $\Delta t$  extremely small. It is even possible for this wave to be completely removed after only a single time step.

The error's dependence on wavelength can be understood by considering how the finite difference grid represents harmonics of different wavelengths. The shortest resolvable wave, with only two data points per wavelength, is poorly represented. As the wavelength increases, the representation improves and approaches the continuous representation as the wavelength tends to infinity.

### 3.4 Time differencing schemes

In this chapter we consider ordinary differential equations with one dependent and one independent variable. Although atmospheric models are essentially always models for solving a complex set of partial differential equations, in some formulations the numerical solution of ordinary differential equations forms an important part of the computational procedure. For instance in spectral models the governing partial differential equations reduce to a set of ordinary differential equations for the expansion coefficients as dependent variables. A set of ordinary differential equations will also be obtained if a Lagrangian method is used, in which the computational points move with the fluid. But, most of all, schemes for solving ordinary differential equations are of interest here since they are often used without modification to construct approximations to the time derivative terms in the governing partial differential equations. Knowledge of the properties of schemes for solving ordinary differential equations will then be used in investigating the properties of more complex schemes for solving the partial differential equations.

With that in mind, we shall here first define some of the schemes that will be interesting to analyze. Then we shall investigate the behavior of numerical solutions obtained when these schemes are used for two specific ordinary differential equations: the oscillation (or frequency) equation, and the friction equation. These equations will serve as prototypes for later extension of the results to advection, gravity-inertia wave, and diffusion processes within the atmospheric primitive equations.

Schemes used for the time derivative terms within the primitive equations are relatively simple, usually of the second and sometimes even only of the first order of accuracy. There are several reasons for this. First, it is a general experience that schemes constructed so as to have a high order of accuracy are mostly not very successful when solving partial differential equations. This is in contrast to the experience with ordinary differential equations, where very accurate schemes, such as the Runge-Kutta method, are extremely rewarding. There is a basic reason for this difference. With an ordinary differential equation, the equation and a single initial condition is all that is required for an exact solution. Thus, the error of the numerical solution is entirely due to the inadequacy of the scheme. With a partial differential equation, the error of the numerical solution is brought about both by the inadequacy of the scheme and by insufficient information about the initial conditions, since they are known only at discrete space points. Thus, an increase in the accuracy of the scheme improves only one of these two components, and the results are not too impressive.

Another reason for not requiring a scheme of high accuracy for approximations to the time derivative terms is that, in order to meet a stability requirement of the type discussed in the preceding sections, it is usually necessary to choose a time step significantly smaller than that required for adequate accuracy. With the time step usually chosen, other errors, for example in the space differencing, are much greater than those due to the time differencing. Thus, computational effort is better spent in reducing these other errors, and not in increasing the accuracy of the time differencing. This, of course, does not mean that it is not necessary to consider carefully the properties of various possible time differencing schemes. Accuracy, is only one important consideration in choosing a scheme.

To define some schemes, we consider the equation

$$\frac{dU}{dt} = f(U, t) \quad U = U(t) \quad (3.19)$$

We divide the time axis into segments of equal length  $\Delta t$ . We shall denote by  $U^{(n)}$  the approximate value of  $U$  at time  $n\Delta t$ . We assume that we know at least the first of the values  $U^{(n)}, U^{(n-1)}, \dots$  and we want to construct a scheme for computation of an approximate value  $U^{(n+1)}$ . These are many possibilities.

### Two-level schemes

These are schemes that relate values of the dependent variable at two time levels:  $n$  and  $n+1$ . Only a two level scheme can be used to advance an integration over the first time step, when just a single initial condition is available. With such a scheme we want to approximate the exact formula

$$U^{(n+1)} = U^{(n)} + \int_{n\Delta t}^{(n+1)\Delta t} f(U, t) dt \quad (3.20)$$

Schemes that don't use an iterative procedure:

**Euler (forward) scheme** This is the scheme

$$U^{n+1} = U^{(n)} + \Delta t \cdot f^{(n)} \quad (3.21)$$

where

$$f^{(n)} = f\left(U^{(n)}, n\Delta t\right)$$

The truncation error of this scheme is  $o(\Delta t)$ . Thus, this is a first order accurate scheme. For the integrand in 3.20 we have here taken a constant value equal to that at the lower boundary of the time interval. Thus, this is an uncentered scheme.

**Backward scheme** We can also take a constant value of  $f$  equal to that at the upper boundary of the time interval. We then obtain

$$U^{n+1} = U^{(n)} + \Delta t f^{(n+1)} \quad (3.22)$$

If, as here, a value of  $f$  depending on  $U^{(n+1)}$  appears in the difference equation, the scheme is called implicit. For an ordinary differential equation, it may be simple to solve such a difference equation for the desired value  $U^{(n+1)}$ . But, for partial differential equations, this will require solving a set of simultaneous equations, with one equation for each of the grid points of the computation region. If a value of  $f$  depending on  $U^{(n+1)}$  does not appear in the difference equation the scheme is called explicit. The truncation error of this scheme is also is  $o(\Delta t)$ .

**Trapezoidal scheme** If we approximate  $f$  in 3.20 by an average of the values at the beginning and the end of the time interval, we obtain the trapezoidal scheme:

$$U^{(n+1)} = U^{(n)} + \frac{1}{2}\Delta t \left( f^{(n)} + f^{(n+1)} \right) \quad (3.23)$$

This is also an implicit scheme. Its truncation error, however, is  $o(\Delta t^2)$ . To increase the accuracy or for other reasons we can also construct iterative schemes. Two level schemes that we will now define are constructed in the same way except an iterative procedure is used to make them explicit.

**Matsuno (or Euler-backward) scheme** With this scheme a step is made first using the Euler scheme; the value of  $U$  obtained for time level  $n + 1$  is then used for an approximation to  $f^{n+1}$ , and this approximate value  $f^{(n+1)}$  is used to make a backward step. Thus,

$$\begin{aligned} U^{*(n+1)} &= U^{(n)} + \Delta t f^{(n)} \\ U^{n+1} &= U^{(n)} + \Delta t f^{*(n+1)} \end{aligned} \quad (3.24)$$

where

$$f^{*(n+1)} = f \left( U^{*(n+1)}, (n+1)\Delta t \right)$$

this is an explicit scheme, of the first order of accuracy.

**Heun scheme** An approximation of the trapezoidal scheme. Thus,

$$\begin{aligned} U^{*(n+1)} &= U^{(n)} + \Delta t f^{(n)} \\ U^{(n+1)} &= U^{(n)} + \frac{1}{2} \Delta t \left( f^{(n)} + f^{*(n+1)} \right) \end{aligned} \quad (3.25)$$

Thus, this is also an explicit scheme. It is of the second order of accuracy.

### Three level schemes

Except at the first step, one can store the value  $U^{(n-1)}$ , and construct schemes taking advantage of this additional information. These are three level schemes. They may approximate the formula

$$U^{(n+1)} = U^{(n-1)} + \int_{(n-1)\Delta t}^{(n+1)\Delta t} f(U, t) dt \quad (3.26)$$

or, they can use the additional value  $U^{(n-1)}$  to make a better approximation to  $f$  in 3.20.

**Leapfrog scheme** The simplest way of making a centered evaluation of the integral in 3.26 is to take for  $f$  a constant value equal to that at the middle of the interval  $2\Delta t$ . This gives the leapfrog scheme

$$U^{(n+1)} = U^{(n-1)} + 2\Delta t \cdot f^{(n)} \quad (3.27)$$

Its truncation error is  $o(\Delta t^2)$ . This is probably the scheme most widely used at present in atmospheric models. It has also been called the "mid-point rule", or "step-over" scheme.

**Adams-Bashforth scheme** The scheme that is usually called the Adams-Bashforth scheme in the atmospheric sciences is, in fact, a simplified version of the original Adams-Bashforth scheme, which is of the fourth order of accuracy. The simplified version is obtained when in 3.20 is approximated by a value obtained at the centre of the interval  $\Delta t$  by a linear extrapolation using values  $f^{(n-1)}$  and  $f^{(n)}$ . This gives

$$U^{(n+1)} = U^{(n)} + \Delta t \left( \frac{3}{2} f^{(n)} - \frac{1}{2} f^{(n-1)} \right) \quad (3.28)$$

This also is a second order accurate scheme. There are many other possibilities. For examples, one can approximate the integral 3.26 using Simpson's rule, that is, by fitting a parabola to the values  $f^{(n-1)}$ ,  $f^{(n)}$  and  $f^{(n+1)}$ . The implicit scheme obtained in this way is called the Milne-Simpson scheme. To illustrate the wealth of possible alternatives we note that in a paper by Young (1968) properties of 13 different schemes have been studied. Furthermore, when we are solving a more complicated partial differential equation, or a system of such equations, time (or space-time) differencing schemes can be constructed which are more complex than those which can be defined using the simple equation 3.19. Such schemes are widely used in atmospheric models, and some of them will be described in later chapters of this publication.

### 3.4.1 Properties of schemes applied to particular equations

The advective equation

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0$$

has a wave solution  $u(x, t) = Re[U(t)e^{ikx}]$  provided that

$$\frac{dU}{dt} + ikc = 0$$

this ordinary differential equation reduces to 3.36 if we substitute  $\omega = -kc$ . The general solution of 3.36 is

$$U(t) = U(0)e^{i\omega t}$$

or, for discrete values  $t = n\Delta t$ :

$$U(n\Delta t) = U(0)e^{in\omega\Delta t} \quad (3.29)$$

Thus, considering the solution in a complex plane, its argument rotates by  $\omega\Delta t$  in each time step and  $\Delta t$  there is no change in amplitude.

The properties of various schemes when applied to 3.36 are conveniently analyzed using the von Neumann method. This method, as we have seen, involves defining a variable  $\lambda$  by

$$U^{(n+1)} = \lambda U^{(n)} \quad (3.30)$$

with

$$\lambda = |\lambda|e^{i\theta}$$

so

$$U^{(n)} = |\lambda|U^{(0)}e^{in\theta} \quad (3.31)$$

with  $\theta$  representing the change in argument (or phase change) of the numerical solution in each time step. Since we know that the amplitude of the true solution does not change, we shall require  $|\lambda| \leq 1$  for stability. It will also be instructive to compare the phase change

unstable	$ \lambda  > 1$
neutral	$ \lambda  = 1$
damping (or dissipative)	$ \lambda  < 1$

Figure 3.3: stability scheme

of the numerical solution per time step,  $\theta$ , with that of the true solution,  $\omega\Delta t$ . The ratio of these changes,  $\frac{\theta}{\omega\Delta t}$ , is the relative phase change of the numerical solution. For accuracy,

Accelerating	$\frac{\theta}{(\omega\Delta t)} > 1$
No effect on phase speed	$\frac{\theta}{(\omega\Delta t)} = 1$
Decelerating	$\frac{\theta}{(\omega\Delta t)} < 1$

Figure 3.4: phase change

therefore, it is desirable to have both the amplification factor and the relative phase speed close to unity. Exceptions to this are so-called “computational modes”, which, as we shall see later, can appear as false solutions superposed on the physical solution. These are solutions that do not approach the true solution as the space and time steps approach zero. If such solutions exist they will each have their own value of the amplification factor. Since they are not an approximation to the true solution, it is desirable to have their amplitudes as small as

possible, that is, to have their amplification factors less than unity.

Stability will require, where  $B$  is a constant

$$|U^{(n)}| = |\lambda|^n |U^{(0)}| < B \quad (3.32)$$

or

$$n \log |\lambda| < \log \frac{B}{|U^{(0)}|} \quad (3.33)$$

$$\log |\lambda| < \log \frac{B \Delta t}{t |U^{(0)}|} \propto \frac{B'}{t} \Delta t$$

we want however that it is valid for finite times

$$\log |\lambda| \leq o(\Delta t)$$

This means that if  $\lambda = \delta$  the deviation must be such that  $\delta \leq o(\Delta t)$  so the condition for stability is:

$$|\lambda| \leq 1 + o(\Delta t) \quad (3.34)$$

this condition will allow for exponential growth, but not faster.

**Generale formula** The three non-iterative two level schemes can be describes by a single finite difference equation:

$$U^{(n+1)} = U^{(n)} + \Delta t (\alpha f^n + \beta f^{(n+1)}) \quad (3.35)$$

with a consistency requirement

$$\alpha + \beta = 1$$

Notice:

- $\alpha = 1, \beta = 1$  for the Euler scheme
- $\alpha = 0, \beta = 1$  for the backward scheme
- $\alpha = 1/2, \beta = 1/2$  for the trapezoidal scheme

**Oscillation equation** For applications in atmospheric models it is of particular interest to consider the case where

$$f = i\omega U$$

Model problem:

$$\begin{aligned} \frac{dU}{dt} &= i\omega U \\ U^{(n)} &\rightarrow U(n\Delta t) \end{aligned} \quad (3.36)$$

The equation 3.36 is known as the oscillation equation or frequency equation. Allowing  $U$  to be complex, 3.36 can be thought of a representing a system of two equations. The parameter  $\omega$  is real, and called the frequency. Applying 3.30 to the oscillation equation gives

$$U^{(n+1)} = U^{(n)} + i\omega \Delta t (\alpha U^n + \beta U^{(n+1)}) \quad (3.37)$$

In order to evaluate  $\lambda$  we must solve this equation for  $U^{(n+1)}$ . Denoting  $p = \omega \Delta t$ :

$$U^{(n+1)} = \frac{1 + i\alpha p}{1 - i\beta p} U^{(n)} \quad (3.38)$$

Therefore

$$\lambda = \frac{1 + i\alpha p}{1 - i\beta p} \quad (3.39)$$

**Euler**  $\alpha = 1 \beta = 0 \lambda = 1 + ip$  so

$$|\lambda| = (1 + p^2)^{1/2}$$

The Euler scheme is always unstable, however for very small time steps  $\Delta t$   $p$  is so small that

$$|\lambda| \approx 1 + \frac{1}{2}p^2$$

The less restrictive von Neumann criterion is satisfied. However, experience shows that an indiscriminate use of the Euler scheme for solution of the atmospheric equations leads to amplification at a quite unacceptable rate.

**Backward**  $\alpha = 0 \beta = 1 \lambda = \frac{1}{1+p^2} 1 + ip$  so

$$|\lambda| = (1 + p^2)^{-1/2}$$

The backward scheme is always stable (no matter what  $\Delta t$ ), however is also dumping, meaning it's reducing the amplitude of the solution. The amount of damping increases as the frequency  $\omega$  increases. This is often considered to be a desirable property of a scheme. For instance, we can think of a system in which a number of frequencies are present at the same time; for example, solving a system of equations of the type 3.36. This situation is similar to that existing in the real atmosphere. It would appear to be necessary to maintain the amplitudes of motions of different frequencies in the correct ratio. However, in numerical integrations, high frequency motions are often excited to unrealistically large amplitudes through errors in the initial data. It may then be desirable to reduce the amplitudes of high frequency motions by a selective damping in the time differencing scheme. In other words, a scheme with frequency dependent damping properties can be used to filter out undesirable high frequency motions.

**Trapezoidal**  $\alpha = 1/2 \beta = 1/2 \lambda = \frac{1}{1+1/4p^2} (1 - 1/4p^2 + ip)$

$$|\lambda| = 1$$

The trapezoidal scheme is, thus, always neutral. The amplitude of the numerical solution remains constant, just as does that of the true solution. It is useful to note that both the implicit schemes considered here were stable no matter how large a value of  $\Delta t$  was chosen.

**Matsuno** The iterative two level schemes can also be described by a single equation in the same way as 3.35. Thus, we write

$$\begin{aligned} U^{(n+1)*} &= U^{(n)} + \Delta t f^{(n)} \\ U^{(n+1)} &= U^{(n)} + \Delta t \left( \alpha f^{(n)} + \beta f^{*(n+1)} \right) \end{aligned} \quad (3.40)$$

that applied to the oscillation equation gives:

$$\begin{aligned} U^{(n+1)*} &= U^{(n)} + i\omega \Delta t U^{(n)} \\ U^{(n+1)} &= U^{(n)} + i\omega \Delta t U^{*(n+1)} \end{aligned} \quad (3.41)$$

eliminating the intermediate values  $U^{(n+1)*}$ :

$$U^{(n+1)} = (1 - \beta p^2 + ip) U^{(n)}$$

thus,

$$\lambda = 1 - \beta p^2 + ip \quad (3.42)$$

Substituting the appropriate values of  $\beta$  for the Matsuno scheme, we obtain

$$\lambda = 1 - p^2 + ip \quad (3.43)$$

and the stability condition (we need to evaluate  $|\lambda|$ ):

$$|\lambda| = (1 - p^2 + p^4)^{1/2} \quad (3.44)$$

the Matsuno scheme is stable if  $|p| < 1$ , so the time step must be

$$\Delta t \leq \frac{1}{|\omega|}$$

This scheme is conditionally stable: the higher the frequency, the more restrictive is the stability condition. When differentiating 3.44 we find that the amplification factor of the Matsuno scheme has a minimum for  $p = 1/\sqrt{2}$ . Therefore, as pointed out by Matsuno (1966a) when dealing with a system with a number of frequencies we can choose a time step so as to have  $0 \leq p \leq 1/\sqrt{2}$  for all the frequencies present, and then, in the same way as backward implicit scheme, this scheme will reduce the relative amplitudes of high frequencies. This technique has recently become very popular for initialization of atmospheric models, where it is used to damp the spurious high frequency noise generated by the assimilation of the observed data. As shown by Matsuno (1966b) higher order accuracy schemes with similar filtering characteristics can be constructed. Figure 3.5 summarizes the results obtained for

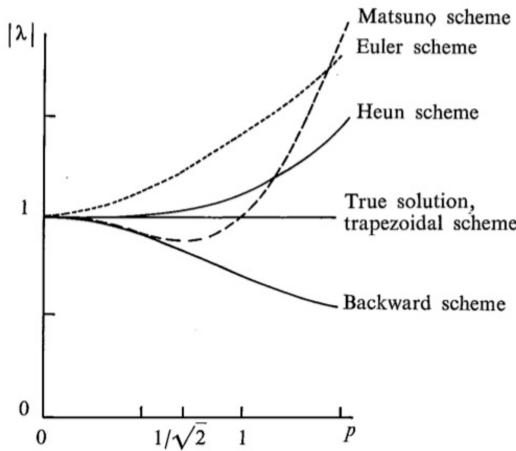


Figure 3.5: The amplification factor as a function of  $p = \omega\Delta t$ , for the five two level schemes and for the true solution

the five schemes considered so far. For all of these schemes the amplification factors were found to be even functions of  $p$ , so the amplification factor curves are shown only for  $p \geq 1$ .

**Leap-Frog scheme (Three level schemes)** The more accurate leap-frog scheme is more delicate.

$$U^{(n+1)} = (U^{(n-1)} + 2i\omega\Delta t U^{(n)}) \quad (3.45)$$

A problem with all three or more level schemes including this is that they require more than one initial condition to start the computation. From a physical standpoint a single initial condition  $U^{(0)}$  should have been sufficient. However, in addition to the physical initial condition, three level schemes require a computational initial condition  $U^{(1)}$ . This value cannot be calculated by a three level scheme, and, therefore, it will usually have to be obtained using one of the two level schemes. According to 3.30 we also have

$$U^{(n)} = \lambda U^{(n-1)} \quad U^{(n+1)} = \lambda^2 U^{(n-1)} \quad (3.46)$$

substituted in 3.45 we have

$$\lambda^2 - 2ip\lambda - 1 = 0$$

that has solutions:

$$\begin{aligned}\lambda_1 &= \sqrt{1 + p^2} + ip \\ \lambda_2 &= -\sqrt{1 + p^2} + ip\end{aligned}$$

Thus, there are two solutions of the form  $U^{(n+1)} = \lambda U^{(n)}$ . This necessarily follows from the fact that we are considering a three level scheme; substitution of 3.46 into the difference equation given by these schemes will always give a second degree equation for  $\lambda$ . In general, an  $m$  level scheme will give  $m-1$  solutions of the form  $U^{(n+1)} = \lambda U^{(n)}$ . A solution of this type corresponding to a single value of  $\lambda$  is called a mode.

Consider now the two values that have been obtained for  $\lambda$ . If a solution of the form  $U^{(n+1)} = \lambda U^{(n)}$  is to represent an approximation to the true solution, then we must have  $\lambda \rightarrow 1$  as  $\Delta t \rightarrow 0$ . For the values  $\lambda_1, \lambda_2$ , as  $p = \omega \Delta t \rightarrow 0$  we do have  $\lambda_1 \rightarrow 1$ , however at the same time  $\lambda_2 \rightarrow -1$ . Solutions like that associated with  $\lambda_1$  are usually called *physical modes* because we are always solving equations describing physical processes. Solutions like that associated with  $\lambda_2$  are not approximations to the true solution, and are called *computational modes*.

- Case  $|p| < 1$ : then  $|\lambda| = 1$  stable
- Case  $|p| = 1$ : then  $|\lambda| = 1$  equal
- Case  $|p| > 1$ : then  $|\lambda| = |p \pm \sqrt{p^2 - 1}|$  unstable

### Stability for dissipation equation

For the dissipation equation

$$\frac{dU}{dt} = -kU, \quad U = U(t), \quad k > 0 \quad (3.47)$$

Again it is easy to justify our interest in this equation. For example, if we define  $U = u + iv$ , it describes the effect of friction proportional to the velocity vector, as is often assumed for motions near the ground. As another example, note that when seeking a solution of the heat transfer, or Fick's diffusion equation in the form of a single harmonic component. The general solution of 3.47 is

$$U(t) = U(0)e^{-kt} \quad (3.48)$$

Thus, both the real and the imaginary part decrease exponentially with time. The properties of schemes applied to 3.47 will again be analyzed using the von Neumann method. As in the previous section, we consider first the non-iterative two level scheme 3.35. Applied to the diffusion equation, it gives:

$$U^{(n+1)} = U^{(n)} - k\Delta t (\alpha U^{(n)} + \beta U^{(n+1)}) \quad (3.49)$$

writing  $K = k\Delta t$  we have

$$U^{(n+1)} = \frac{1 - \alpha K}{1 + \beta K} U^{(n)} \quad (3.50)$$

and

$$\lambda = \frac{1 - \alpha K}{1 + \beta K} \quad (3.51)$$

For the two level schemes we have:

- Euler ( $\alpha = 1, \beta = 0$ ):  $\lambda = |1 - K| \rightarrow 0 < K \leq 2$  stable
- Backward ( $\alpha = 0, \beta = 1$ ):  $\lambda = \frac{1}{1+K} \rightarrow K > 0$  stable
- Trapezoidal ( $\alpha = 1/2, \beta = 1/2$ ):  $\lambda = \frac{1-1/2K}{1+1/2K} \rightarrow K > 0$  stable

while for the leap-frog scheme:

$$\lambda^2 + 2K\lambda - 1 = 0$$

giving the solutions

$$\begin{aligned}\lambda_1 &= -K + \sqrt{1 + K^2} \\ \lambda_2 &= -K - \sqrt{1 + K^2}\end{aligned}\tag{3.52}$$

As  $k \rightarrow 0$ ,  $\lambda_1 \rightarrow 1$ , while  $\lambda_2 \rightarrow -1$  thus, the solution associated with  $\lambda_1$  again represents the physical mode, and that associated with  $\lambda_2$  the computational mode. For  $K > 0$ , that is for the normal case of a forward integration in time, this result always in a computational mode  $\lambda_2$  that is always unstable. It changes sign from time step to time step, and its magnitude increases. As before, we cannot hope to eliminate the computational mode completely. Because we can't eliminate completely the computational mode the leapfrog cannot be used for the dissipation equation. The same analysis applied to the friction equation will give a different result. The Euler scheme is stable and the leapfrog scheme is instead is always unstable.

### 3.4.2 A combination of schemes

A natural question to ask at this point is what can we do if, for example, the equation contains both the oscillation and the dissipation term, that is

$$\frac{dU}{dt} = i\omega U - kU\tag{3.53}$$

Here we might like to use the leapfrog scheme because of the oscillation term  $i\omega U$ , but we know that it cannot be used for the dissipative term  $-kU$ . In this and similar situations we can use different schemes for the different terms; for example, we might use the leapfrog scheme for the oscillation term and the forward scheme for the dissipative term. We then obtain:

$$U^{(n+1)} = U^{(n-1)} + 2\Delta t \left( i\omega U^{(n)} - kU^{(n-1)} \right)\tag{3.54}$$

## 3.5 The Advection Equation

We now consider differential equations with one dependent and two or three independent variables, that is, partial differential equations. More specifically, we shall consider various simplified forms of the advection equation, describing advection of a dependent variable. This is considered in practice to be the most important part of the atmospheric governing equations.

We have already discussed the one-dimensional linear advection equation to some extent in the introductory chapter. We shall organize the analysis here so as first to continue considering problems associated with the simplest one-dimensional linear form of the advection equation, and then to proceed to problems introduced by more complex forms of the advection equation. Consider the equation

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0\tag{3.55}$$

here  $u = u(x, f)$  is a function of two independent variables: the independent variable  $x$  will represent a space variable, and, thus, 3.55 will be called the one-dimensional linear advection equation. As seen earlier, its general solution is

$$u = f(x - ct)$$

where  $f$  is an arbitrary function. One of the finite difference schemes for 3.55 is the forward and upstream scheme. If the space derivative (i.e. discretizing only " $x$ ") is approximated by a centered finite difference quotient using values at the two nearest points, we obtain for the time derivative:

$$\frac{\partial u_j}{\partial t} = -c \frac{u_{j+1} - u_{j-1}}{2\Delta x} \quad (3.56)$$

The subscript here, as before, denotes the distance from the origin in space increments; that is,  $x = j\Delta x$ . A number of schemes for the numerical solution of 3.55 can now be constructed because we can approximate the time derivative in 3.56 by one of the methods studied in the preceding chapter. For example, when the time derivative is approximated using the leapfrog scheme, we obtain

$$\frac{u_j^{n+1} - u_j^{n-1}}{2\Delta t} = -c \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} \quad (3.57)$$

as one of many possible consistent schemes for the numerical solution of 3.55.

A possible solution of 3.55 will be of the form

$$u_j = \operatorname{Re} [U(t) e^{ikj\Delta x}] \quad (3.58)$$

going back to the oscillation equation:

$$\frac{dU}{dt} = i \left( -\frac{c}{\Delta t} \sin k\Delta x \right) U \quad (3.59)$$

denoting

$$p = -c \frac{\Delta t}{\Delta x} \sin k\Delta x$$

Now, if we approximate 3.59 using one of the time differencing schemes, the same finite difference equation is obtained as when we apply that scheme to 3.56 and then substitute the wave solution 3.58. Hence, properties of finite difference schemes derived from 3.56 can be inferred from the results of Section Time differencing schemes, where the frequency to is given by  $p = \omega\Delta t$ . As an example, if 3.59 is approximated using the leap-frog scheme, we obtain:

$$U^{(n+1)} = U^{(n-1)} + 2 \left( -c \frac{\Delta t}{\Delta x} \sin k\Delta x \right) U^{(n)} \quad (3.60)$$

with

$$p = -c \frac{\Delta t}{\Delta x} \sin k\Delta x \quad (3.61)$$

We obtain the same finite difference equation 3.60 by first applying the leapfrog scheme to 3.56 giving 3.57 and then substituting 3.58 into 3.57. Thus, properties of 3.57 can be inferred from the known properties of the leapfrog scheme applied to the oscillation equation.

Let us look at some of conclusions obtained in this way. For stability of the leapfrog scheme it was required that the condition  $|p| \leq 1$  occurring. Thus, we have to satisfy

$$\left| c \frac{\Delta t}{\Delta x} \sin k\Delta x \right| \leq 1 \quad (3.62)$$

for any admissible  $k$ . Since  $|\sin k\Delta x|$  does reach the maximum value of unity in the admissible range of  $k$ , we obtain the stability condition

$$|c| \frac{\Delta t}{\Delta x} \leq 1 \quad (3.63)$$

This critetion (already mentioned) shows that stability cannot simply be achieved by reduction of the time and space increments. Rather, it is necessary to reduce the ratio of these

increments  $\frac{\Delta x}{\Delta t}$  to obtain stability. This condition was first found by Courant, Friedrichs and Lewy (1928) and therefore, is usually referred to as the Courant-Friedrichs-Lowy, or CFL, stability criterion.

It is instructive to note that the maximum value of  $|p|$ , that is, the minimum stability, is associated with the wave with  $k\Delta x = \frac{\pi}{2}$ . This is the component with wave length  $4\Delta x$ , twice the shortest resolvable wave length  $2\Delta x$ .

Let us use now another scheme to approximate the time derivative in 3.56: the Matsuno scheme. First the approximate values  $u_j^{(n+1)*}$  using the forward scheme, that is,

$$\frac{u_j^{(n+1)*} - u_j^n}{\Delta t} + c \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} = 0 \quad (3.64)$$

then, these approximate values are used in the backward scheme, that is,

$$\frac{u_j^{(n+1)} - u_j^n}{\Delta t} + c \frac{u_{j+1}^{(n+1)*} - u_{j-1}^{(n+1)*}}{2\Delta x} = 0 \quad (3.65)$$

eliminating the intermediate step, i.e. the approximate values  $u^{(n+1)*}$ , from this equation, by substituting values given by 3.64, with the subscript  $j$  replaced by  $j+1$  and then by  $j-1$ .

$$\begin{aligned} \frac{u_{j+1}^{(n+1)*} - u_{j+1}^n}{\Delta t} + c \frac{u_{j+2}^n - u_j^n}{2\Delta x} &= 0 \\ \frac{u_{j-1}^{(n+1)*} - u_{j-1}^n}{\Delta t} + c \frac{u_j^n - u_{j-2}^n}{2\Delta x} &= 0 \end{aligned}$$

CFL criterion. In this way we obtain:

$$\frac{u_j^{(n+1)} - u_j^n}{\Delta t} = -c \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x} + c^2 \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{2\Delta x^2} \quad (3.66)$$

Without the last term, this represents the finite difference equation obtained when the forward scheme is used for the time derivative in 3.56. The third term approaches zero as  $\Delta x, \Delta t \rightarrow 0$ , and 3.66 is therefore also a consistent scheme for the advection equation. On the other hand, for a fixed  $\Delta t$  this term approaches  $c^2 \Delta t \left( \frac{\partial^2 u}{\partial x^2} \right)$  as  $\Delta x \rightarrow 0$ . It is therefore of the same form as a finite difference approximation to a Fick's diffusion term, and it has a damping effect. This damping effect, however, is dependent on the wave length. As the third term is calculated over an interval of  $4\Delta x$ , the maximum damping occurs for a wave with wave length of  $4\Delta x$ . There is no damping of the shortest resolvable wave with wave length  $2\Delta x$ . Even if a damping effect were desirable when solving the advection equation, we would not want this particular dependence on wave length. Thus, the Matsuno scheme does not appear suitable for solving the advection equation.

It is convenient to include here one more example of the use of the energy method for testing stability. In addition to being applicable also to nonlinear equations, it can be used to study the effect of boundary conditions on stability. We will use the energy method here to test the stability of a group of schemes that can be used for solving 3.56. The last term can be decomposed as

$$\frac{u_{j+1}^n - 2U_j^n + u_{j-1}^n}{2\Delta x^2} = \left( \frac{u_{j+2}^n - u_j^n}{2\Delta x} \right) - \left( \frac{u_j^n - u_{j-2}^n}{2\Delta x} \right) \frac{1}{2\Delta x} \approx \frac{\partial^2 u}{\partial x^2} \quad (3.67)$$

The scheme is adding a dissipative term.

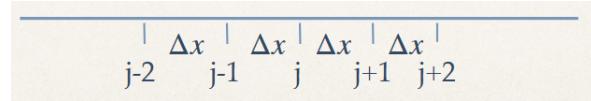


Figure 3.6: steps

### Non linear advection

Consider now the non-linear advection equation

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} \quad (3.68)$$

consider the effect of the non linear term on a single wave  $u = \sin kx$ , a function  $u(x)$  which can be represented by values at grid points. When performed on finite differences it cannot resolve wave length shorter than  $2\Delta x$ , or wave numbers greater than  $k_{max} = \pi/\Delta x$ ,  $k < k_{max}$ . For the wave the non-linear term will look like:

$$u \frac{\partial u}{\partial x} = k \sin kx \cos kx = \frac{1}{2}k \sin 2kx \quad (3.69)$$

If  $k$  is in the range  $\frac{1}{2}k_{max} < k \leq k_{max}$  the result of the multiplication will be beyond the capability of the grid to represent. It cannot, therefore, be properly reproduced in a finite difference calculation. To gain some insight into what happens in such a situation, consider a wave for which  $k > K_{max}$ . For example, let  $L_t = \frac{4\Delta x}{3}$ . A wave of that wave length is shown by the full line in 4.6.2. Knowing only the values at grid points we will not be able to distinguish this wave from the one shown by the dashed line. Thus, with the convention adopted earlier which assumes that the longest waves are present, we will make an error. This is called *aliasing error*. The waves with wave number  $k_u > k_{max}$  will be interpreted as longer

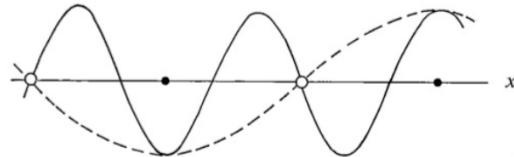


Figure 3.7: A wave of wave length  $\frac{4\Delta x}{3}$ , misrepresented by the finite difference grid as a wave of wave length  $4\Delta x$ .

waves according to  $k^* = 2k_{max} - k_u$ . In a real numerical solution, higher wave numbers will be continuously be created by the non linear interactions and so the amplitude of wave numbers close to  $k_{max}$  will continue to increase.

Now consider the consequences of aliasing errors in a numerical integration. An atmospheric variable, as a function of space co-ordinates, can be thought of as consisting of a series of harmonic components. It is useful to consider the “energy” of these components, that is, their contribution to the mean square value of the variable considered as a function of wave number.

This is the *spectrum* of the “energy”. For example, if the variables are velocity components, this function is the kinetic *energy spectrum*. This spectrum describes the relative importance of features of different scales in the field of the variable. Now, experience shows that the spectrum of atmospheric variables does not change much with time. On synoptic maps we

do not have situations where small scale features are dominant on one day, and absent on the next. Accordingly, spectra of atmospheric variables also do not change much in their general shape. The energy of a particular component can, of course, change, but the characteristic shape of the spectrum as a whole is fairly constant. For example, a zonal spectrum of the eastward velocity component in middle latitudes typically has a maximum for wave numbers 4 to 7, that is, 4 to 7 wave lengths along a latitude circle, with the energy tapering off rather rapidly as the wave number increases beyond about 10. Thus, there is very little energy in wave numbers of the order of the maximum wave numbers that can be resolved by finite difference grids used in atmospheric models.

In a finite difference integration, in addition to these relatively small physical changes, the shape of a spectrum is subject to changes due to aliasing errors. If we have a spectrum of the shape just described, and consider the representation of various combinations  $k_1 + k_2$  that are greater than  $k_{max}$ , we see that most of the energy of such combinations will belong to components with wave numbers not much greater than  $k_{max}$ . Thus, due to aliasing errors a spurious energy inflow is expected at wave numbers that are not much less than  $k_{max}$ , and, in time, the energy of these components can be expected to grow beyond physically acceptable limits. Experience shows that, if no precautionary measures are taken, this can indeed happen, and even cause a catastrophic end to the integration. The phenomenon is due to the nonlinear terms of the equations, and, therefore, is called the *nonlinear instability*. Nonlinear instability was first encountered by Norman Phillips (1956) in his famous work that laid the foundation for the numerical modeling of the atmospheric general circulation. Starting from an atmosphere at rest, he integrated the vorticity equation for a simulated time of the order of 30 days. The calculation then came to an end due to an explosive increase in the total energy of the system, associated with an appearance of elongated shapes in the vorticity field. Phillips initially believed that the breakdown was due to excessive truncation errors, and he later repeated the experiment using space and time steps both reduced to about half of their previous values. This must have greatly reduced the truncation errors, but the catastrophic increase in total energy still happened at about the same time.

In a later paper Phillips (1959) gave the interpretation of nonlinear instability similar to what has been presented here, but for the nondivergentvorticity equation. For a test of this explanation Phillips again repeated his experiment, but after every two hours of simulated time he performed a harmonic analysis of the vorticity fields, and eliminated all components with  $k > \frac{1}{2}k_{max}$ . If there are no components with  $k > \frac{1}{2}k_{max}$  the advection term cannot produce waves with  $k > k_{max}$ . We expect that it will be some time before the amplitudes of the eliminated waves are built up again to an appreciable extent. This filtering procedure eliminated the appearance of the spurious increase in energy, thereby confirming this explanation of the instability.

### 3.6 Energy and numerical methods

*References for the following section can be found in the book at chapter 4.2: Finite differencing in two dimensions*

Some issues with nonlinear instability:

- The nonlinear instability shows the issues with controlling errors that make the amplitude grow indefinitely
- On the other end, this issues shows that there are conservation property of the basic equations that must be reflected in the numerical scheme.
- This remark led to the development or the requirement that numerical scheme must be devised that have the same conservation properties of the original equation

- To illustrate this point we will use the simplified homogeneous flow system that we have introduced previously.

The selection of appropriate spatial finite difference formulations to use in the horizontal and vertical directions varies widely depending on which properties are being modeled or are considered important. Most climate models are not intended primarily for predicting the exact location of daily weather events with great accuracy but are directed more toward the simulation of the long-term mean and variance states of the climate system. Since the models frequently run for long periods of time, the numerical schemes used should approximately maintain many of the integral constraints that are in the original differential equations, such as conservation of mass, energy and momentum. Many climate models have employed horizontal finite difference methods that maintain these features.

Consider an incompressible fluid with a variable height, and assume that the horizontal velocity components,  $u$  and  $v$  are independent of height, and that the vertical velocity  $w$  equals 0 at the bottom of the fluid. The set of momentum equations for a homogeneous flow are the following:

$$\begin{aligned} h &= 0 \\ \frac{\partial u}{\partial t} &= -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - g \frac{\partial \eta}{\partial x} & (3.70) \\ \frac{\partial v}{\partial t} &= -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - g \frac{\partial \eta}{\partial y} \\ \frac{\partial \eta}{\partial t} &= -\frac{\partial}{\partial x} u(\eta + H_0) - \frac{\partial}{\partial y} v(\eta + H_0) \end{aligned}$$

Considering the total height  $H = H_0 + \eta$  and multiplying the first two equations by  $H$  we obtain:

$$H \frac{\partial u}{\partial t} = -uH \frac{\partial u}{\partial x} - vH \frac{\partial u}{\partial y} + fHv - gH \frac{\partial \eta}{\partial x} \quad (1)$$

$$H \frac{\partial v}{\partial t} = -uH \frac{\partial v}{\partial x} - vH \frac{\partial v}{\partial y} - fHu - gH \frac{\partial \eta}{\partial y} \quad (2)$$

$$\frac{\partial H}{\partial t} = -\frac{\partial}{\partial x} uH - \frac{\partial}{\partial y} vH \quad (3)$$

Using the relations:

$$H \frac{\partial u}{\partial t} = \frac{\partial}{\partial t}(uH) - u \frac{\partial H}{\partial t} = -\frac{\partial}{\partial x}(uuH) + u \frac{\partial}{\partial x}(uH) - \frac{\partial}{\partial x}(uvH) + u \frac{\partial}{\partial x}(vH) \quad (3.71)$$

We can rewrite the equations of motion as:

$$\begin{aligned} \frac{\partial uH}{\partial t} &= -\frac{\partial uuH}{\partial x} - \frac{\partial uvH}{\partial y} + fHv - g \frac{\partial \eta}{\partial x} & (3.72) \\ \frac{\partial vH}{\partial t} &= -\frac{\partial uvH}{\partial x} - \frac{\partial vvH}{\partial y} - fHu - gH \frac{\partial \eta}{\partial y} \\ \frac{\partial H}{\partial t} &= -\frac{\partial}{\partial x} uH - \frac{\partial}{\partial y} vH \end{aligned}$$

In this form the mass equation makes mass conservation immediately visible, let us now consider energy. Kinetic energy is defined as:

$$KE = \frac{1}{2}H(u^2 + v^2)$$

and the potential energy:

$$PE = \frac{1}{2}gH^2$$

now, by multiplying eq.(1) by  $Hu$  and (3) by  $u^2/2$  and adding, then eq. (2) by  $Hv$  and (3) by  $v^2/2$  and adding, using the product derivative rule as before, we can get:

$$\frac{\partial}{\partial t} \left( H \frac{1}{2} u^2 \right) + \frac{\partial}{\partial x} \left( Hu \frac{1}{2} u^2 \right) + \frac{\partial}{\partial y} \left( Hv \frac{1}{2} u^2 \right) - f H vu + g Hu \frac{\partial H}{\partial x} \quad (3.73)$$

$$\frac{\partial}{\partial t} \left( H \frac{1}{2} v^2 \right) + \frac{\partial}{\partial x} \left( Hu \frac{1}{2} v^2 \right) + \frac{\partial}{\partial y} \left( Hv \frac{1}{2} v^2 \right) - f H vu + g Hv \frac{\partial H}{\partial x} \quad (3.74)$$

$$(3.75)$$

multiplying the continuity equation (3) by  $gH$  and rearranging we get:

$$\frac{\partial}{\partial t} \left( g \frac{1}{2} H^2 \right) + \frac{\partial}{\partial x} (g H^2 u) + \frac{\partial}{\partial y} (g H^2 v) - gH \left[ u \frac{\partial H}{\partial x} + v \frac{\partial H}{\partial y} \right] = 0 \quad (3.76)$$

The total energy is defined as:

$$TE = KE + PE = \frac{1}{2} H(u^2 + v^2) + \frac{1}{2} gH^2$$

hence,

$$\frac{\partial}{\partial t} TE = \frac{\partial}{\partial x} \left( \frac{Hu}{2}(u^2 + v^2) + gH^2 \right) + \frac{\partial}{\partial y} \left( \frac{Hv}{2}(u^2 + v^2) + gH^2 \right) \quad (3.77)$$

Integrating over the domain assuming periodic boundary conditions, we get the result for the time variation of the total energy. The integration of the boundary terms is zero and so we get the final result showing the conservation of the total energy.

### 3.6.1 Conservation in numerical schemes

Let's go back to the simple problem

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} = -\frac{\partial}{\partial x} \left( \frac{u^2}{2} \right) \quad (3.78)$$

In numerical form (any schemes would do) for a centered difference scheme:

$$\frac{\partial u}{\partial t} \approx -\frac{u_{j+1}^2 - u_{j-1}^2}{4\Delta x} \quad (3.79)$$

The energy will be the sum over the grid point values.

$$\int \frac{\partial}{\partial t} u^2 = \int u \frac{\partial}{\partial t} u \approx \sum_{j=1}^{N-1} \frac{u_{j+1}^2 - u_{j-1}^2}{4\Delta x} u_j \quad (3.80)$$

If we expand the sum we notice that terms do not compensate each other so the scheme does not conserve energy. The problem arise from the terms that are mixed in a way that does not compensate. However, in the original equation there are compensations, so maybe we can find ways of writing the equations in a equivalent way.

$$\frac{\partial u}{\partial t} = -\frac{1}{3} \left[ u \frac{\partial u}{\partial x} + \frac{\partial}{\partial x} u^2 \right]$$

using the same centered scheme:

$$\frac{\partial u}{\partial t} \approx -\frac{1}{3} \left[ u_j \frac{u_{j+1} - u_{j-1}}{2\Delta x} + \frac{u_{j+1}^2 - u_{j-1}^2}{2\Delta x} \right]$$

will result in a scheme that conserves energy. So, by using alternative but equivalent formulations schemes with desirable conservation properties can be found.

We have seen how find a different thing can be used to solve one-dimensional problems. But the real world is of course three dimensional and we have to understand how to deal with that in order to attack realistic cases. Therefore it is important that we understand how we can apply these methods to more dimensions. To address the problem in more dimensions, let's first introduce a key feature of numerical modeling.

### 3.6.2 The Arakawa Jacobian

The Arakawa Jacobian is central to numerical modeling in geophysical fluid dynamics, particularly for ensuring the conservation and stability of numerical simulations of vorticity and enstrophy (a measure of the vorticity squared). This concept is foundational in finite-difference schemes used to model fluid flows, such as weather patterns or ocean currents.

A conceptually elegant approach for dealing with the nonlinear instability problem has been suggested and developed by Arakawa (1966, 1972). His idea is that it is better, if possible, to use schemes for the advection terms that are not only free of the nonlinear computational instability but also free of the spurious inflow of energy to these short waves, instead of artificially suppressing their amplitudes. The amplitudes of the shortest waves in atmospheric models are small initially, and they will remain small if a false generation of these short waves is avoided. Arakawa has shown that it is possible to construct such schemes, and that they are obtained when care is taken to conserve in the finite difference form some integral properties of the original differential equations.

When the Arakawa conservation schemes are used there is no need for an artificial dissipation in the advection process. This enables the statistical properties of the schemes to be maintained under advection, a feature especially useful in general circulation studies.

Here are the key points and essential concepts regarding the Arakawa Jacobian and its application in two finite-difference schemes. The Arakawa Jacobian is a numerical method used to approximate the Jacobian term in the vorticity equation of fluid dynamics, which is the nonlinear term representing the interaction of velocity fields. This term is crucial in the conservation of key quantities like energy and enstrophy, both of which need to be carefully handled to avoid numerical artifacts and to accurately represent physical processes in models.

A major advantage of the Arakawa scheme is its ability to eliminate nonlinear instability, which arises due to the accumulation of energy at small scales caused by aliasing. By carefully constructing the Jacobian, the scheme conserves both energy and enstrophy, preventing the unphysical growth of energy in smaller scales.

There are many ways of constructing finite difference approximations to the Jacobian. We can use any of the three equivalent analytic expressions. We shall consider only approximations of the second order of accuracy. With the simplest centered space differencing, we require values of  $p, q$  from a box of nine adjacent grid points to evaluate the Jacobian.

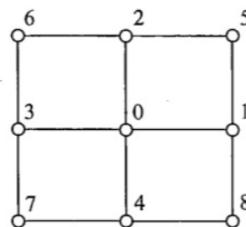


Figure 3.8: Stencil used to define approximations to Jacobian

The construction of the Arakawa Jacobian involves combining multiple second-order approximations, denoted as  $J^{++}, J^{x+}, J^{+x}$ , with specific weights chosen to conserve energy,

enstrophy, and vorticity. The superscripts + and  $x$  denote the positions of the points from which values of  $p$  and  $q$ , respectively, are used to form the approximation. Each of the approximations  $J^{++}$ ,  $J^{+x}$  and  $J^{x+}$  is consistent and of the second order of accuracy. A more general approximation can now be formed as a linear combination of these three, that is

$$J(p, q) = \alpha J^{++} + \beta J^{+x} + \gamma J^{x+}$$

with  $\alpha + \beta + \gamma = 1$ . To illustrate that let's use the vorticity equation in the non-divergent form:

$$\frac{\partial}{\partial t} \nabla^2 \psi + J(\psi, \nabla^2 \psi + f_0 + \beta y) = 0 \quad (3.81)$$

The nonlinear instability was generated by the nonlinear advection terms that are represented here by the Jacobian operator  $J(\psi, \nabla^2 \psi)$ . So the term that we have to treat carefully is the Jacobian. The Arakawa Jacobian, that has been used by essentially every climate model using finite differences, is

$$J_A = \frac{1}{3} (J^{++} + J^{+x} + J^{x+})$$

By eliminating the need for artificial dissipation, the scheme preserves the statistical properties of advection, making it particularly valuable for general circulation and climate studies. The Arakawa scheme also controls the computational energy cascade by conserving the average wave number in the advection terms of the nondivergent flow, addressing nonlinear instability more effectively than simpler energy-conserving approximations.

Interestingly, the Arakawa Jacobian has been demonstrated to be a special case of the finite element method, where a variational formulation minimizes errors through interpolation. However, while the scheme ensures conservation of energy and enstrophy, these properties alone do not guarantee the accuracy of the solution. Thus, the Arakawa Jacobian remains a powerful tool for controlling numerical instability, even though solution accuracy must be verified through additional means.

Thus, it is not only a conservation of energy, as has sometimes incorrectly been implied. For example, an approximation can easily be constructed for the nonlinear term of the one-dimensional advection equation 3.68 which would conserve the kinetic energy. Using such an approximation, however, would not prevent nonlinear instability in the way that the Arakawa scheme does. The Arakawa procedure does not have a one-dimensional analogue, as the non-divergent vorticity equation 3.81 is not nonlinear when applied to a one-dimensional problem.

### 3.6.3 Gravity waves

In this chapter we consider the equations describing the horizontal propagation of gravity and gravity-inertia waves. Mathematically, this means that we will be dealing with a system of two or three partial differential equations of the first order. Thus, we will now have two or three dependent variables. The system of equations will always be equivalent to a single differential equation of a higher order. This equation can be obtained from the system by elimination of dependent variables.

We first put this problem in perspective. Arakawa, (Arakawa, 1970), has stated that there are two main problems in finite difference integrations of the atmospheric governing equations. One is a proper simulation of the geostrophic adjustment process. Through this process the atmosphere establishes a characteristic quasi-non divergent state, mostly as a result of the dispersion of the gravity-inertia waves. The associated computational considerations will be discussed in this chapter. The second problem is the prediction or simulation of the large-scale quasi-non divergent flow after it has been established. Here the horizontal advection is the dominating mechanism.

Extensive study of the problems in integrations of the gravity-inertia wave equations began in atmospheric modeling much later than studies of the advection problem. After Richardson's (1922) first unsuccessful numerical integration of the complete primitive equations, the successful result of Charney, Fjortoft and von Neumann (1950) was largely due to the exclusion of gravity-inertia waves from their equations by using the geostrophic approximation in the vorticity equation. The governing equations with the gravity-inertia waves excluded, are customarily called the filtered equations. They bypass the geostrophic adjustment problem. The filtered equations were used almost exclusively in the first decade of numerical forecasting research.

Efforts to improve the performance of numerical models led to a desire to include the non-geostrophic effects. This is very difficult to do within the modified system of equations. Thus, starting with the first successful experiments by Hinkelmann (1959), modelers came back to using the primitive equations. Except for special purposes, the primitive equations are used almost exclusively in atmospheric models today. They are generally considered superior for both research and operational applications (e.g. Sawyer, 1972). The speed of propagation of the gravity and gravity-inertia waves, and their sensitivity to various numerical errors mean that their treatment requires especially careful consideration.

The more general case of one dimension the gravity waves:

$$\begin{aligned}\frac{\partial u}{\partial t} &= -g \frac{\partial H}{\partial x} \\ \frac{\partial H}{\partial t} &= -H_0 \frac{\partial u}{\partial x}\end{aligned}$$

will generate the differencing:

$$\frac{\partial u_j}{\partial t} = -g \frac{H_{j+1} - H_{j-1}}{2\Delta x} \quad (3.82)$$

$$\frac{\partial H_j}{\partial t} = -H_0 \frac{u_{j+1} - u_{j-1}}{2\Delta x} \quad (3.83)$$

with

$$c^* = \pm \sqrt{gH_0} \frac{\sin k\Delta x}{k\Delta x} \quad (3.84)$$

This phase speed is a function of wave number, and, thus, we see that the space differencing again results in computational dispersion. The formula 3.84 is the same as the one obtained in the preceding chapter when considering the advection equation. Therefore, both the phase speed and the group velocity depend on the wave number. The phase speed decreases as the wave length decreases, and the wave with wave length  $2\Delta x$  is stationary.

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 3.82 is concerned, however, the underlined



Figure 3.9: A grid with two dependent variables that are both carried at every grid points.

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 "sub-grids", with the solution on one of these sub-grids being completely decoupled from the other. Thus, it would be better to calculate only one of these solutions, that is, to use a grid as shown below in 3.10 Such a grid, with variables carried at alternate points in space, is

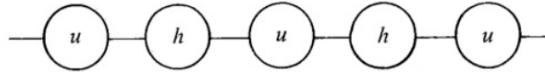


Figure 3.10: A grid with two dependent variables that are carried at alternate grid points.

called a staggered grid. The computation time needed to solve 3.82 on the grid is reduced by a factor of two, and the truncation error is the same. Furthermore, the waves with  $k\Delta x > \frac{\pi}{2}$  have been eliminated, and these are just the waves associated with large phase speed errors and negative group velocities.

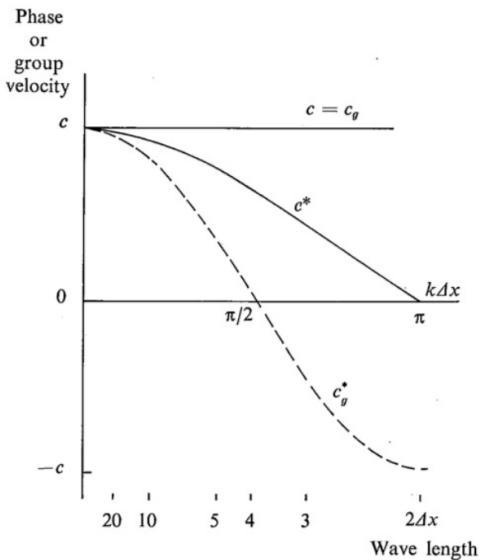


Figure 3.11: Phase speed and group velocity, in the case of the linear advection equation,  $c$  and  $c_g$ , and in the case of the corresponding differential-difference equation with second-order centered space differencing,  $c^*$  and  $c_g^*$

Notice: for the exact advection equation 3.7 both individual waves and wave packets, that is, places where superposition of waves results in a maximum amplitude of a group of neighboring wave numbers, propagate at the same constant velocity  $c = c_g$ . Introduction of the centered space finite difference quotient both makes the phase speed and the group velocity decrease as the wave number increases. The error is particularly great for the shortest resolvable wave lengths; waves with wave lengths less than  $4\Delta x$  even have a negative group velocity. This means that wave packets made up of these waves propagate in the direction opposite to the advection velocity and opposite to the direction of propagation of individual waves. Thus, when using such a staggered grid, the phase speed and group velocity diagram shown in 3.11 is reduced to its left half, covering waves with wave lengths of up to  $4\Delta x$  only. This is a tremendous improvement. If we wish to have waves with wave lengths between  $4\Delta x$  and  $2\Delta x$  in our calculation we can reduce the grid length by a factor of two and perform a much more accurate integration, using the same amount of computation time than with a grid that is not staggered.

### 3.6.4 Two dimension

The system of equations for two-dimensional gravity waves is given by:

$$\begin{aligned}\frac{\partial u}{\partial t} &= -g \frac{\partial H}{\partial x} \\ \frac{\partial v}{\partial t} &= -g \frac{\partial H}{\partial y} \\ \frac{\partial H}{\partial t} &= -H_0 \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)\end{aligned}\quad (3.85)$$

with  $\omega^2 = gH_0(k^2 + l^2)$  the wave frequency. With two dimensions and three dependent variables, a large number of spatial arrangements of the variables are possible. For the present we consider the three rectangular arrangements shown below. The identifying letters (A), (E) and (C) are chosen so as to conform with the symbols used by Winninghoff and Arakawa (Arakawa, 1972). Grid arrangements significantly influence the accuracy and efficiency of

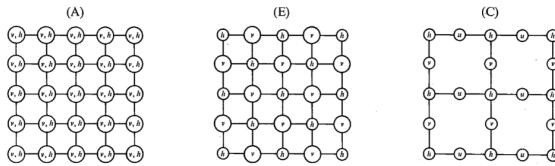


Figure 3.12: Three types of lattice considered for the finite difference solution of 3.85

finite difference solutions. Among the considered grids (A, E, and C), grid (C) offers the most accurate phase speed for gravity waves due to its spatial configuration, but it complicates calculations of Coriolis forces because velocity components are not collocated. The admissible regions of wave numbers in the wave number plane can be found by considering the shortest resolvable wave lengths. Note that with lattice (E) the lines joining the nearest points with the same variable make an angle of  $\frac{\pi}{4}$  with the grid lines while with the other two lattices these lines are along the grid lines. Grid (E), with its staggered arrangement, provides computational advantages over grid (A), reducing dispersion errors and requiring less computation time while achieving comparable results. However, it is prone to issues with two-grid-interval waves arising from decoupling of subgrids, which can lead to spurious stationary solutions.

Two-grid-interval waves are caused by the decoupling of solutions on (C) subgrids in the (E) lattice. These manifest as spurious stationary waves or low-frequency inertia oscillations. They are triggered by artificial boundary conditions, latent heat release, or orographic forcing. This issue highlights the importance of addressing subgrid interactions in numerical models.

When Coriolis terms are included in the equations:

$$\begin{aligned}\frac{\partial u}{\partial t} &= -g \frac{\partial H}{\partial x} + fv \\ \frac{\partial v}{\partial t} &= -g \frac{\partial H}{\partial y} - fu \\ \frac{\partial H}{\partial t} &= -H_0 \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)\end{aligned}\quad (3.86)$$

their inclusion complicates grid selection. Grid (C), ideal for pure gravity waves, struggles with Coriolis terms. Grids (B) and (E) perform similarly but exhibit dispersion issues for the shortest waves. The geostrophic adjustment process, governed by these equations, disperses high-frequency waves until the motion approaches geostrophic balance. Accurate simulation of this process requires careful selection of grid arrangements and numerical methods.

It is not obvious how we should analyse various arrangements of variables. Our primary concern here is to consider 3.86 as part of the complete system of primitive equations. We are interested in large-scale motions, otherwise we would not be including the Coriolis terms.

On the large scale, the primitive equations admit two distinct types of motion : low-frequency, quasi-geostrophic and quasi-nondivergent flow and high-frequency gravity-inertia waves. Gravity-inertia waves are continually excited in the atmosphere. However, as they are dispersive, a local accumulation of wave energy disperses with time. This process is known as geostrophic adjustment; the remaining motion is in approximate geostrophic balance and changes only slowly in time. In this chapter we are concerned with the correct simulation of this process, which is essentially governed by the gravity-inertia wave equations 3.86.

We consider five ways of distributing the dependent variables in space, shown in 3.9. We denote by  $d$  the shortest distance between neighbouring points carrying the same dependent variable. In the figure  $d$  is the same for each of the five lattices; thus, all the lattices have the same number of dependent variables per unit area. The computation time needed for an integration on each of the lattices will be about the same; the properties of the solution obtained, though, will differ because of the effect of the space arrangement of variables.

Using the subscripts shown in the figure, we define the centered space differencing operator by

$$\delta_x \alpha = \frac{1}{d'} (\alpha_{i+1/2,j} - \alpha_{-1/2,j}) \quad (3.87)$$

this rotation is applicable to all the lattices. Here  $d'$  is the distance between the points between which the finite difference is taken. Thus, for lattices (A) through (D)  $d'$  is equal to the grid size  $d$ , and for the lattice (E) it is equal to  $\sqrt{2d}$ .

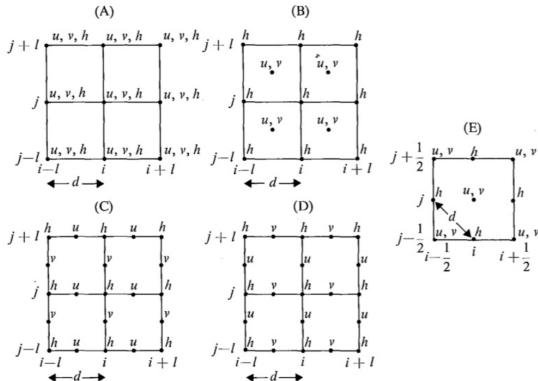


Figure 3.13: Five types of lattice considered for the finite difference solution of 3.86

The phase speed function,

$$\left(\frac{\omega}{f}\right)^2 = 1 + \frac{gH}{f^2} k^2 \quad (3.88)$$

shows that frequency increases monotonically with wave number, ensuring the group velocity  $\frac{\partial v}{\partial k}$  is never zero. This prevents local wave energy accumulation during geostrophic adjustment. The non-dimensional frequency  $\frac{\omega}{f}$  depends on  $kd$  and  $\frac{\lambda}{d}$ , where the radius of deformation is

$$\lambda = \sqrt{\frac{gH}{f}}$$

never equal to zero.

Dissipative numerical schemes can be useful in filtering high-frequency unphysical waves during the geostrophic adjustment process. These schemes prevent excessive wave amplitudes caused by artificial boundary conditions or abrupt grid changes. However, dissipation must be

applied selectively to avoid distorting physical processes. For gravity-inertia wave terms, dissipation can accelerate the adjustment process by damping unphysical high-frequency waves. This approach is acceptable when only the final geostrophic state is of interest, but it should be avoided if the high-frequency waves themselves are the focus of study.

Grid (C) provides the closest approximation to the exact solution, avoiding false group velocities, while grids (B) and (E) exhibit issues with false low frequencies and stationary solutions. Grids (D) and (A) have poor dispersion properties and are unsuitable for accurate simulations.

In conclusion, grid (C) is the best choice for simulating geostrophic adjustment due to its superior dispersion properties and minimal spurious wave generation. Staggered grids (B) and (E) are viable alternatives but require methods to address spurious low-frequency waves. Proper dissipation, applied selectively, can enhance numerical stability without compromising physical accuracy.

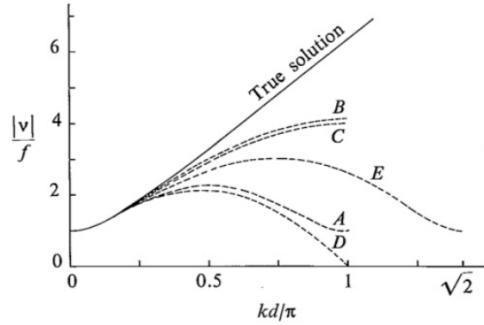


Figure 3.14: The function  $\frac{\omega}{f}$  given by 3.88. The wavelength of the shortest resolvable wave along the  $x$  axis is  $sd$  for lattices (A) through (D), and  $\sqrt{2d}$  for the lattice (E). Thus, we have to consider the range  $0 < kd \leq \pi$  for lattices (A) through (D), and the range  $0 < kd \leq \sqrt{2\pi}$  for the lattice (E).

If we consider the one-dimensional case in which the dependent variables are constant along the lines , we obtain results for these two lattices. In general, we define the coordinate system  $x'$ ,  $y'$  by rotating the  $x$ , $y$  in the positive direction through an angle of  $\pi/4$ , and then, using some velocity transformations we change from variables  $u$ ,  $v$  to new dependent variables  $u'$ ,  $v'$ . We find that the dispersion properties of the lattices (B) and (E) can be considered equivalent. A gravity-inertia wave in one of these lattices has phase speed and dispersion properties identical to those of the same wave with its front rotated through an angle of  $\pi/4$  in the other lattice.

Considering the two dimensional case, the values of  $\frac{\omega}{f}$  that are obtained in the two-dimensional case for the true solution and those using lattices (B) and (C) are shown in 3.15. The diagram for lattice (E) can be obtained by a counterclockwise rotation of the (B) lattice diagram. In the two-dimensional case, lattice (C) provides a significantly better

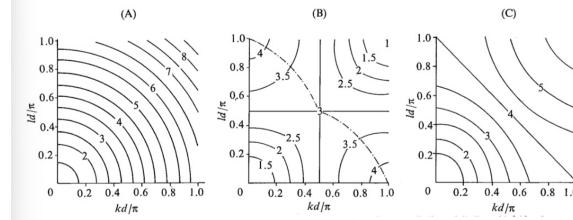


Figure 3.15: The function  $\omega/f$  for the true solution

approximation to the exact solution compared to lattices (B) or (E). In the (B) lattice diagram,

a dot-dashed line indicates a maximum value of  $\omega/f$  for a given ratio  $l/f$ , but such a feature is absent in both the (C) lattice diagram and the exact solution. This absence highlights the advantage of lattice (C), as it avoids waves with group velocities of the wrong sign. However, the effectiveness of lattice (C) depends on the parameter  $\lambda/d$ , where  $\lambda$  is the radius of deformation. If  $\lambda/d$  approaches 1 or less, as might occur in conditions of weak stability, the advantages of lattice (C) diminish. Fortunately, for typical grid sizes in atmospheric models, this is not an issue, leading Arakawa and Lamb (1976) to conclude that lattice (C) is the best choice for simulating geostrophic adjustment processes.

Lattices (B) and (E), on the other hand, struggle with false low frequencies in the shortest waves. Specifically, the two-grid-interval wave, which was stationary as a pure gravity wave, now behaves like a pure inertia oscillation. This issue arises from the decoupling of gravity wave solutions on the two complementary (C)-type subgrids that make up these lattices. Strategies to address these problems are discussed in later sections.

Dissipation in numerical schemes should only be used selectively. For advection, dissipative schemes are generally unnecessary and can lead to energy cascading falsely into short waves. However, short waves may still appear due to boundary reflections, sudden grid changes, or rapidly changing coefficients. In such cases, limited dissipation can help mitigate these effects.

For gravity-inertia waves, dissipation can be helpful in damping high-frequency unphysical waves, speeding up the geostrophic adjustment process. This adjustment relies on wave dispersion, not dissipation, but selective dissipation can approximate the result. If only the final balanced state is important, dissipation is acceptable. However, if high-frequency waves are the focus, dissipation should be avoided to preserve accuracy.

## 3.7 Spectral Methods

Spectral methods are an approach to discretization, the main idea is to express my basic variables in terms of other functions. In some cases the use of finite differences is a simple and direct method to solve differential equations. However this is not the only method. There are other ways in which one can solve differential equations numerically using a special approach that is based on special functions.

### Properties of spectral methods

- Derivative becomes multiplication, i.e. algebraic operations.
- Within the limits of the truncation the numerical derivative is exact.
- There is no aliasing.
- The smallest wavelength in the truncation corresponds to the minimum scale that can be resolved.

The main example of this approach is the usage of trigonometric functions within the context of the more general theory of Fourier analysis. Consider the Oscillation equation:

$$\frac{\partial^2 \Phi}{\partial t^2} = c^2 \frac{\partial^2 \Phi}{\partial x^2} \quad (3.89)$$

where  $\Phi$  is the deviation from the rest position. This equation could also express the vibration modes of a string of length L fixed at the extremities. That mathematically is expressed by the fact that the boundary conditions at the extremes are zero. Defining  $\Psi$  the unknown variable:

$$\Phi(x, t) = \Psi(x)e^{i\omega t} \quad (3.90)$$

with  $e^{i\omega t} = \cos \omega t + i \sin \omega t$  also

$$\frac{\partial^2 \Psi}{\partial x^2} + k^2 \Psi = 0 \quad (3.91)$$

with  $k = \frac{\omega}{c}$  The solution is either

$$\Psi(x) = A \sin(kx + \delta)$$

or

$$\Psi(x) = A \cos(kx + \delta)$$

The boundary conditions for fixed extremities ( $\Psi(0) = \Psi(L) = 0$ ) exclude the cosine solution, so the general solution is

$$\Psi_n(x) = A_n \sin(nkx) \quad (3.92)$$

this is not the only one function: I'll have an integer number of wavefunctions that satisfy this condition meaning that  $k$  cannot be arbitrary, i.e. the boundary conditions are satisfied only for integer values of  $k$

$$k = \frac{\pi}{L}, \frac{2\pi}{L}, \dots, \frac{n\pi}{L}$$

(only sin functions that have an integer number of cycles can be a solution.) We'll get infinite number of solutions that satisfy the boundary conditions. These are known as eigenfunctions (or vectors)  $\Psi_n(x)$  of the wave equation for the string, and they are linear transformations acting on finite vector (they are of the same number as the dimension of the matrix). The most important property of the eigenfunction is the **orthogonality**:

$$\int_0^L \Psi_n(x) \Psi_m(x) dx = C_n \delta_{nm} \quad (3.93)$$

with

$$\delta_{mn} = \begin{cases} 1, & \text{if } n = m \\ 0, & \text{if } n \neq m \end{cases}$$

The eigenfunctions can always be normalized so that the constants  $C$  becomes 1. It is easy to show that in this case the normalized eigenfunctions are:

$$\Psi_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L} \quad (3.94)$$

The most important property of the eigenfunction is the orthogonality, in fact: For  $n = m$ :

$$\int_0^L \sin^2 \frac{n\pi x}{L} dx = \frac{1}{2} \int_0^L \left[ 1 - \cos \frac{2n\pi x}{L} \right] dx = \frac{x}{2} \Big|_0^L - \frac{L}{4\pi n} \sin \frac{2n\pi x}{L} \Big|_0^L = \frac{L}{2}$$

for  $n \neq m$ :

$$\begin{aligned} \int_0^L \sin \frac{n\pi x}{L} \sin \frac{m\pi x}{L} dx &= \frac{1}{2} \int_0^L \left[ \cos \frac{(n-m)\pi x}{L} - \cos \frac{(n+m)\pi x}{L} \right] dx = \\ &= \frac{1}{2} \frac{L}{n-m} \sin \frac{(n-m)\pi x}{L} \Big|_0^L - \frac{1}{2} \frac{L}{n+m} \sin \frac{(n+m)\pi x}{L} \Big|_0^L = 0 \end{aligned}$$

The orthogonality can be used to describe every function  $f(x)$ , meaning that I can generalize the concept managing to write functions as sum of eigenfunctions:

$$f(x) = \sum_n a_n \Psi_n(x)$$

How can I find the coefficients? Multiply both sides by another eigenfunction  $\Psi_k(x)$  and integrate:

$$\int_0^L f(x)\Psi_k(x)dx = \sum_n a_n \int_0^L \Psi_n(x)\Psi_k(x)dx = \sum_n a_n \delta_{kn} = a_n$$

so it seems that every function can be expressed as a sum of countable terms. This sum is over integer index,  $f(x)$  is continuous. I should find out its behavior in every point, this can be described by a slighter lower order of infinity.

**Fourier theory** Every function can be expressed in a set of eigenfunction, expanded over a series of the trigonometric functions.

$$f(x) \approx A_0 + \sum_{k=1}^n A_k \cos kx + \sum_{k=1}^n B_k \sin kx \quad (3.95)$$

where the A's and B's are computed from orthogonality relations as we've seen before.

$$\begin{aligned} A_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)dx \\ A_\pi &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx dx \\ B_\pi &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx dx \end{aligned}$$

We have already seen the importance of Euler formula  $e^{i\theta} = \cos \theta + i \sin \theta$ , it is useful to see the form of the Fourier series in terms of the complex exponential, substituting the expression for the coefficients in the un-truncated representation of  $f(x)$ .

$$f(x) = \sum_{k=-\infty}^{+\infty} c_k e^{ikx} \quad (3.96)$$

with

$$c_k = \frac{1}{2\pi} \int f(\xi) e^{-ik\xi} d\xi \quad (3.97)$$

This approximation works fine if everything is smoothed, but if there are jumps it increases amplitude over them, this is known as Gibbs phenomenon. The Gibbs phenomenon happens when we use a Fourier series to represent a function with a sudden jump (discontinuity). Near the jump, the Fourier series creates wavy overshoots and undershoots. As we add more terms to the series, the first overshoot stays at about 9% of the jump size, and the waves don't completely go away. Instead, they get squeezed closer to the jump, so their overall effect becomes smaller, and the total area of the waves approaches zero. At the jump point, the Fourier series gives the average of the function's both side limits toward the point.

This new representation of functions has some problems, even though manages to solve problems with higher order: accuracy. The eigenfunction expansion, the Fourier series, allows us to find a new representation of functions. What we have found is an alternative but equivalent representation of a function that has all the same properties of the original function. The question now arise if we can use these alternative form to find a way to drive a numerical solution that is somewhat better or cheaper or easier.. Applying this into a set of ordinary equations, let's for instance take the advection equation:

$$\frac{\partial u}{\partial t} = -c \frac{\partial u}{\partial x}$$

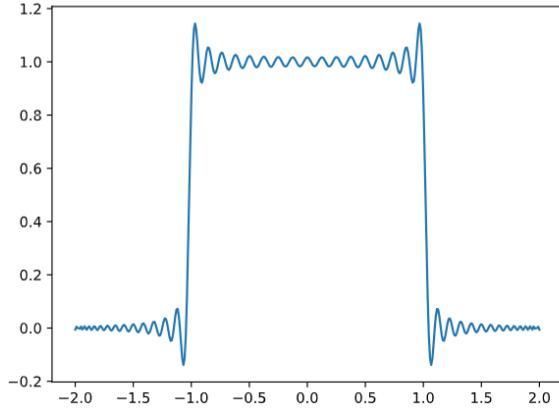


Figure 3.16: Representation of Gibbs phenomenon.

and let's use a Fourier representation using the first  $N$  components:

$$u(x) = \sum_{k=1}^N u_k(t) e^{ikx} \quad (3.98)$$

with the initial condition:

$$u(x, 0) = \sum_{k=1}^N u_k(0) e^{ikx} \quad (3.99)$$

now substituting, multiplying by  $e^{ijkx}$  and integrating over  $x$ , with

$$\int e^{ikx} e^{ijx} dx = \delta_{ij}$$

We get back exactly, in the linear case

$$\frac{du_k}{dt} = -cku_k(t) \quad (3.100)$$

### 3.7.1 Numerical spectral solution (nonlinear)

Consider now the nonlinear advection equation:

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} \quad (3.101)$$

Let's try a spectral solution substituting eq.3.98 and eq.3.99,

$$\sum_k \frac{d}{dt} u_k(t) e^{ikx} = - \sum_j u_j(t) e^{ijx} \sum_k \frac{d}{dx} u_k(t) e^{ikx} = - \sum_{k,j} iku_j(t) u_k(t) e^{ijx} e^{ikx}$$

projecting on another eigenfunction and integrating:

$$\int \sum_k \frac{du_k}{dt} e^{ikx} e^{inx} dx = - \int \sum_{k,j} ik u_j(t) u_k(t) e^{ijx} e^{ikx} e^{inx} dx$$

the integrals yield:

$$\begin{aligned} \int e^{ikx} e^{inx} dx &= \delta_{kn} \\ \int e^{ikx} e^{ikx} e^{inx} dx &= c_{ikn} \end{aligned}$$

so the spectral system looks like

$$\frac{du_k}{dt} = - \sum_{j,n} iku_j(t)u_k(t)c_{kjn} \quad (3.102)$$

These are now nonlinear ordinary differential equation and we have to compute the  $c_{kjn}$  coefficients, also known as *interaction coefficients*. The equations are not clearly separated, every spectral component is depending on its evolution on every other component  $u_j$ . The interaction coefficients are defined as:

$$c_{knj} = \delta_{k,j+n} = \begin{cases} 1, & \text{if } k = j + n \\ 0, & \text{otherwise} \end{cases} \quad (3.103)$$

The method is very convenient and it has a lot of good properties ( by the way, quadratic properties are conserved, within the truncation limit). However the number of coefficients grows very rapidly with the truncation limit as  $N^3$ . Furthermore, every particular non linear term will generate its own set of interaction coefficients, so there will be many sets of c's to compute and store. These days a typical climate model will have truncation of the order of 100, so for each term  $10^6$  coefficient will have to be computed.

### 3.8 Spherical frame

*The following section refers to the slides of prof. Navarra and his discussion, in the book this topic is faced differently.*

We will use the shallow water equation, but in order to have a realistic case, we need to put it on the sphere. We start from the usual equation of the homogeneous flow, also known as shallow water equation on the sphere. For convenience we introduce the geopotential  $\Phi = gH$ , representing the gravitational potential energy per unit mass related to height in the atmosphere.

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - \frac{\partial \Phi}{\partial x} \quad (3.104)$$

$$\frac{\partial v}{\partial t} = -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - \frac{\partial \Phi}{\partial y} \quad (3.105)$$

$$\frac{\partial \Phi}{\partial t} = -\frac{\partial}{\partial x}u\Phi - \frac{\partial}{\partial y}v\Phi \quad (3.106)$$

Let's now consider the momentum equation and equation of continuity in vector form. Using a spherical rather than a  $(x, y)$  grid,  $u$  and  $v$  become longitudinal and latitudinal components of wind. It is convenient to use the representation of scalar fields on the globe, so write velocities  $U = u \cos \phi$  and  $V = v \cos \phi$  in terms of the latitude  $\phi$ , so we can write the system in vector form using the horizontal wind  $\mathbf{V} = iU + jV$  ( $i$  is the longitude component east-west, while  $j$  is the latitudinal north-south):

$$\begin{aligned} \frac{\partial \mathbf{V}}{\partial t} + (\mathbf{V} \cdot \nabla) \mathbf{V} &= -f \mathbf{k} \times \mathbf{V} - \nabla \Phi & (3.107) \\ \frac{\partial \Phi}{\partial t} + (\mathbf{V} \cdot \nabla) \Phi &= -\Phi \nabla \cdot \mathbf{V} \end{aligned}$$

$\mathbf{k}$  is the vertical component. In the geopotential equation,  $(\mathbf{V} \cdot \nabla) \Phi$  is the advection term at the geopotential, while the right side of the equation represents the divergence. Using the identity

$$(\mathbf{V} \cdot \nabla) \mathbf{V} = \nabla \left( \frac{\mathbf{V} \cdot \mathbf{V}}{2} \right) + \zeta \mathbf{k} \times \mathbf{V}$$

that allows us to modify non-linearity of the advection  $(\mathbf{V} \cdot \nabla) \mathbf{V}$ , we can re-write the first equation of 3.107 as

$$\frac{\partial \mathbf{V}}{\partial t} = -(\zeta + f)\mathbf{k} \times \mathbf{V} - \nabla \left( \Phi + \frac{\mathbf{V} \cdot \mathbf{V}}{2} \right) \quad (3.108)$$

On the left we have the tendency (rate of change of velocities); on the right:

1. the changes in advection in zonal components is actually driven by vorticity and  $f$  (Coriolis) coupled with meridional motion.
2. the rate of geopotential generating tendency to the currents. Different gradients in the speed will generate acceleration or deceleration according to the gradient of the current, this term is coming from the advection. The advection is operating with vorticity in a cross way, also kinetic energy  $(\mathbf{V} \cdot \mathbf{V})$  is operating.

Using the definitions of vorticity (curl of the velocity field) and divergence (divergence of the velocity field). Vorticity  $\zeta$  represents the rotational or swirling motion of the fluid, while the divergence  $D$  represents the expansion or contraction of the fluid, i.e. the rate at which the fluid is spreading out or converging.

$$\zeta = \mathbf{k} \cdot (\nabla \times \mathbf{V}) \quad (3.109)$$

$$D = \nabla \cdot \mathbf{V} \quad (3.110)$$

We can then find the expressions for the time rate change of vorticity and divergence:

$$\frac{\partial \zeta}{\partial t} = -\nabla \cdot (\zeta + f)\mathbf{V} \quad (3.111)$$

because, from the definition and the eq.3.108, the geopotential and kinetic term goes to zero as the curl of the gradient is zero. In this equation  $\zeta$  is the relative vorticity, while  $f$  is the planetary vorticity. And for the divergence we find:

$$\frac{\partial D}{\partial t} = \mathbf{k} \cdot \nabla \times (\zeta + f)\mathbf{V} - \nabla^2 \left( \Phi + \frac{\mathbf{V} \cdot \mathbf{V}}{2} \right) \quad (3.112)$$

It is also useful to divide the geopotential from its global constant mean,  $\bar{\Phi}$ ,  $\Phi = \bar{\Phi} - \Phi'$ , the continuity equation becomes:

$$\frac{\partial \Phi'}{\partial t} = \frac{\partial \Phi}{\partial t} = -(\mathbf{V} \cdot \nabla)(\bar{\Phi} + \Phi') - (\bar{\Phi} - \Phi')D = -[\mathbf{V} \cdot \nabla \Phi' + \Phi' D] - \bar{\Phi}D = -\nabla \cdot \Phi' \mathbf{V} - \bar{\Phi}D$$

The advection will kill the mean, so we'll have only its action on the deviation.

**Why decompose using vorticity and divergence?** By separating the rotational (vorticity) and compressional (divergence) components, you can better understand the dynamics of the flow field, such as areas dominated by rotation (e.g., cyclones, anticyclones) or regions of convergence or divergence (e.g., updrafts, downdrafts). In meteorology and oceanography, the large-scale velocity field is often approximated using vorticity ( $\zeta$ ) and divergence ( $D$ ) because these quantities are central to the equations of motion, making them more directly tied to the dynamics than  $\mathbf{V}$  itself.

Any vector field (like  $\mathbf{V}$ ) also can be decomposed into two components using Helmholtz theorem:

- Irrotational (divergence-driven) derived from a scalar potential  $\chi$
- Solenoidal (vorticity-driven) derived from a streamfunction  $\psi$

The idea is starting from the velocity field  $\mathbf{V}$ , getting the vorticity and divergence equations, solve for the streamfunction  $\psi$  and the scalar velocity potential  $\chi$  using  $\zeta$  and  $D$  (see below), and reconstruct the velocity field using Helmholtz theorem.

### 3.8.1 The Helmholtz theorem

Helmholtz showed that any vector field tangent to the surface of a sphere can be split in terms of two scalar functions, one rotational (velocity) and another nonrotational (divergence), in the following way:

$$\mathbf{V} = \mathbf{k} \times \nabla\psi + \nabla\chi \quad (3.113)$$

where  $\mathbf{k}$  is the vector normal to the surface of the sphere.

- The **streamfunction** is a scalar function used to describe two-dimensional (or axisymmetric three-dimensional) incompressible flow. It simplifies the analysis of fluid motion by providing a way to represent the velocity field such that the incompressibility condition is automatically satisfied. It describes the rotational part of the flow.

By the definitions 3.109, we can write through the Helmholtz theorem:

$$\zeta = \nabla^2\psi \quad (3.114)$$

$$D = \nabla^2\chi \quad (3.115)$$

We could also derive these definitions from the description of  $D$  and  $\zeta$  in spherical coordinates:

$$\begin{aligned} D &= \frac{1}{a \cos \phi} \left( \frac{\partial u}{\partial \lambda} + \frac{\partial v \cos \phi}{\partial \phi} \right) = \frac{1}{a \cos^2 \phi} \left( \frac{\partial U}{\partial \lambda} + \cos \phi \frac{\partial V}{\partial \phi} \right) \\ \zeta &= \frac{1}{a \cos \phi} \left( \frac{\partial v}{\partial \lambda} - \frac{\partial u \cos \phi}{\partial \phi} \right) = \frac{1}{a \cos^2 \phi} \left( \frac{\partial V}{\partial \lambda} - \cos \phi \frac{\partial U}{\partial \phi} \right) \end{aligned}$$

using the expression for  $U$  and  $V$ , for instance in the divergence:

$$D = \frac{1}{a \cos^2 \phi} \left( \frac{\partial^2 \chi}{\partial \lambda^2} + \cos \phi \frac{\partial}{\partial \phi} \left( \cos \phi \frac{\partial \chi}{\partial \phi} \right) \right) = \nabla^2\chi \quad (3.116)$$

so we find that also on the sphere:

$$\zeta = \nabla^2\psi \quad (3.117)$$

and

$$D = \nabla^2\chi \quad (3.118)$$

Because of Helmholtz theorem we can write the components in terms of  $\psi$  and  $\chi$  as:

$$U = -\frac{\cos \phi}{a} \frac{\partial \psi}{\partial \phi} + \frac{1}{a} \frac{\partial \chi}{\partial \lambda} \quad (3.119)$$

$$V = \frac{\cos \phi}{a} \frac{\partial \chi}{\partial \phi} + \frac{1}{a} \frac{\partial \psi}{\partial \lambda} \quad (3.120)$$

And we can find equations for vorticity and divergence rates depending on these new scalar fields, from 3.111 and 3.112:

$$\frac{\partial}{\partial t}(\nabla^2\psi) = -\frac{1}{a \cos^2 \phi} \left[ \frac{\partial}{\partial \lambda}(U \nabla^2\psi) + \cos \phi \frac{\partial}{\partial \phi}(V \nabla^2\psi) \right] - 2\Omega \left( \sin \phi \nabla^2\chi + \frac{V}{a} \right) \quad (3.121)$$

$$\frac{\partial}{\partial t}(\nabla^2\chi) = +\frac{1}{a \cos^2 \phi} \left[ \frac{\partial}{\partial \lambda}(V \nabla^2\psi) - \cos \phi \frac{\partial}{\partial \phi}(U \nabla^2\psi) \right] + 2\Omega \left( \sin \phi \nabla^2\psi - \frac{U}{a} \right) - \nabla^2 \left( \frac{U^2 + V^2}{2 \cos^2 \phi} + \Phi' \right) \quad (3.122)$$

together with:

$$\frac{\partial \Phi'}{\partial t} = -\frac{1}{a \cos^2 \phi} \left[ \frac{\partial}{\partial \lambda}(U \Phi') + \cos \phi \frac{\partial}{\partial \phi}(V \Phi') \right] - \bar{\Phi} D \quad (3.123)$$

make the system complete and the three equations above are three prediction equations for the three unknown variables  $\psi$ ,  $\chi$  and  $\Phi'$ . The systems of equations is more complex than the straightforward use of  $u$ ,  $v$  and  $h$ , but  $\psi$  and  $\chi$  fields are scalar functions that can be expressed easily as spherical harmonics. The transform technique allows for relatively simple means of solving the equations. The essence of this technique is to transform from spectral space quantities to grid point space quantities, then perform nonlinear products at each grid point, and to reverse quantities from grid point quantities back to spectral quantities. The problems in fact are the nonlinearity terms present in the equations, such as:

$$U\nabla^2\psi, \quad V\nabla^2\psi, \quad U\Phi', \quad V\Phi' \quad \text{and} \quad (U^2 + V^2)/2$$

In order to do this, the fields of  $\psi$ ,  $\chi$  and  $\Phi'$  have to be represented as truncated series of spherical harmonics, as described in the next section.

### 3.8.2 Spherical Harmonics

The Laplacian in spherical coordinates gives rise to Laplace equation: eq.3.116=0, setting it up as a search for the eigenfunctions/eigenvalues requires to solve a Poisson equation with periodic boundary conditions in longitude and zero at the poles.

$$\frac{1}{a \cos^2 \phi} \left( \frac{\partial^2 Y}{\partial \lambda^2} + \cos \phi \frac{\partial}{\partial \phi} \left( \cos \phi \frac{\partial Y}{\partial \phi} \right) \right) = \sigma Y \quad (3.124)$$

the solution can be found for discrete values of two indices  $(m, l)$ . The boundary conditions, as in the case of the string, force solution only for discrete set of values. Because the equation is in two dimension there will be two families of functions, labeled by the the indices  $(l, m)$ :

$$Y_l^m = P_l^m(\sin \phi) e^{im\lambda} \quad (3.125)$$

We can see how we can derive the spherical harmonics by noticing that we can look for solutions of the Laplace equation as  $Y = P(\phi)Q(\lambda)$ , then

$$\frac{\partial^2 P(\phi)Q(\lambda)}{\partial \lambda^2} + \cos \phi \frac{\partial}{\partial \phi} \left( \cos \phi \frac{\partial P(\phi)Q(\lambda)}{\partial \phi} \right) = \sigma a \cos^2 \phi P(\phi)Q(\lambda)$$

The two parts are separated in variables so a solution exist only if they are equal to some constant:

$$\frac{1}{Q} \frac{\partial^2 Q(\lambda)}{\partial \lambda^2} = -m^2$$

$$\frac{1}{P(\phi)} \cos \phi \frac{\partial}{\partial \phi} \left( \cos \phi \frac{\partial P(\phi)}{\partial \phi} \right) - \sigma a \cos^2 \phi = m^2$$

The periodic boundary condition on longitude require  $m$  to be real and discrete, so the solution to the  $Q$  equation are longitudinal waves  $e^{\pm im\lambda}$ . The regularity conditions at the poles (essentially conditions for the solutions to be well-behaved) requires that  $\sigma$  be of the form  $l(l+1)$  with  $l \geq |m|$ . Then a change of variables ( $u = \sin \phi$ ) will put the  $P$  equation in the form of the associated Legendre equation whose solutions are the associated Legendre polynomials  $P_l^m(\sin \phi)$ :

$$P_l^m(\sin \phi) = (1 - \sin^2 \phi)^{m/2} \frac{d^m}{d(\sin \phi)^m} P_l(\sin \phi) \quad (3.126)$$

where the  $P_l(\sin \phi)$  are the Legendre polynomials. For  $m = 0$ , the associated Legendre polynomials reduce to the Legendre polynomials, the first few associated Legendre polynomials

with  $m > 0$  are

$$\begin{aligned} P_1^1(\sin \phi) &= (1 - \sin^2 \phi)^{1/2} \\ P_2^1(\sin \phi) &= 3 \sin \phi (1 - \sin^2 \phi)^{1/2} \\ P_2^2(\sin \phi) &= 3(1 - \sin^2 \phi) \end{aligned}$$

they are an orthogonal in variable  $\phi$ , and the full spherical harmonics are orthogonal in both indices:

$$Y_l^m = P_l^m(\sin \phi) e^{im\lambda} \quad (3.127)$$

they are an orthogonal system, with respect both indices, so

$$\int_{-\pi/2}^{\pi/2} \cos \phi d\phi \int_0^{2\pi} Y_l^m(\phi, \lambda) \bar{Y}_i^j(\phi, \lambda) d\lambda = \delta_{mj} \delta_{ij} \quad (3.128)$$

with the normalization:

$$Y_l^m = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\sin \phi) e^{im\lambda}$$

There are useful recursion relation for associated Legendre polynomial and for the spherical

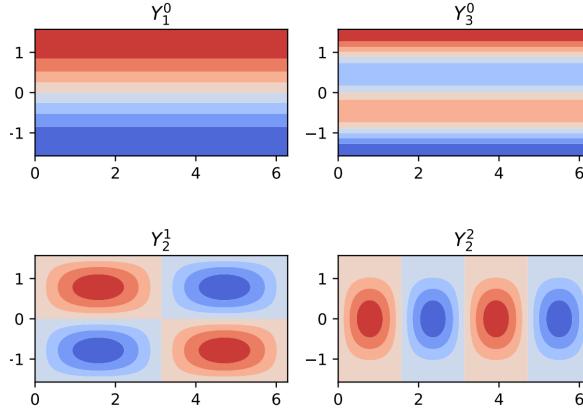


Figure 3.17: Spherical harmonics

harmonics

$$\epsilon_{l+1}^m P_{l+1}^m = \sin \phi P_l^m - \epsilon_l^m P_{l-1}^m$$

and

$$(1 - \sin^2 \phi) \frac{d}{d(\sin \phi)} Y_l^m = -l \epsilon_{l+1}^m Y_{l+1}^m + (l+1) \epsilon_l^m Y_{l-1}^m \quad (3.129)$$

with

$$\epsilon_l^m = \left( \frac{(l^2 - m^2)}{(4l^2 - 1)} \right)^{1/2}$$

Note that the relation 3.129 is effectively transforming the derivation of a spherical harmonics into a difference of other harmonics.

### 3.8.3 Spectral Transforms

Having now a set of functions that are eigenfunctions of the Laplacian on the sphere is then natural that we look for representations of the physical variables as an expansion in terms of spherical harmonics.

$$\begin{aligned}
\psi &= a^2 \sum_{m=-M}^M \sum_{l=|m|}^{|m|+J} \psi_l^m Y_l^m & (1) \\
\chi &= a^2 \sum_{m=-M}^M \sum_{l=|m|}^{|m|+J} \chi_l^m Y_l^m & (2) \\
\Phi &= a^2 \sum_{m=-M}^M \sum_{l=|m|}^{|m|+J} \Phi_l^m Y_l^m & (3) \\
\nabla^2 \psi &= a^2 \sum_{m=-M}^M \sum_{l=|m|}^{|m|+J} \psi_l^m \nabla^2 Y_l^m = a^2 \sum_{m=-M}^M \sum_{l=|m|}^{|m|+J} \left( -\frac{l(l+1)}{a^2} \right) \psi_l^m Y_l^m
\end{aligned}$$

Figure 3.18: spectral representations

Horizontal wind components will have a similar expansion. The expansion for  $U$  and for  $V$  have an extra component as they are derived from stream function and velocity potential. We can use the recurrence formulas of the spherical harmonics to find that

$$\epsilon_{l+1}^m P_{l+1}^m = \sin \phi P_l^m - \epsilon_l^m P_{l-1}^m$$

and

$$\begin{aligned}
U_l^m &= (l-1)\epsilon_l^m \psi_{l-1}^m - (l+2)\epsilon_{l+1}^m \psi_{l+1}^m + im\chi_l^m \\
V_l^m &= -(l-1)\epsilon_l^m \chi_{l-1}^m + (l+2)\epsilon_{l+1}^m \chi_{l+1}^m + im\psi_l^m
\end{aligned}$$

with  $\epsilon_l^m$  defined as before. These expression must be inserted into the equation to obtain ordinary equations for the spherical harmonics coefficients.

Recalling the equation of motion in terms of vorticity and divergence 3.121 and 3.123, the linear terms will transform in linear terms for the spherical coefficients, and the laplacians will also be easily done by the multiplication by the eigenvalues. We will need to treat the nonlinear terms  $UV^2\psi$ ,  $V\nabla^2\psi$ ,  $U\Phi'$ ,  $V\Phi'$ ,  $(U^2 + V^2)/2$ . However, we don't really need all the details of these non linear terms, we only need the total impact on the tendency. In this case, we need only the spectral coefficient of  $A(\psi, \chi)$ ,  $B(\psi, \chi)$ ,  $C(\psi, \chi)$  and  $E(\psi, \chi)$  non linear terms of the equation of motion. With the spectral coefficient of the non linear terms, we can discretize in time by choosing a time scheme, for example the leap-frog scheme.

Rewrite as

$$\begin{aligned}
\frac{\partial \nabla^2 \psi}{\partial t} &= A(\psi, \chi) - 2\Omega \left( \sin \phi \nabla^2 \chi + \frac{V}{a} \right) \\
\frac{\partial \nabla^2 \chi}{\partial t} &= B(\psi, \chi) + 2\Omega \left( \sin \phi \nabla^2 \psi - \frac{U}{a} \right) + E(\psi, \chi) - \nabla^2 \Phi' \\
\frac{\partial \Phi'}{\partial t} &= C(\psi, \chi) - \bar{\Phi} D
\end{aligned}$$

Where the A,B,C,E terms are all nonlinear terms, then the equation for the spectral coefficients could be written as

$$\begin{aligned}
-\frac{l(l+1)}{a^2} \frac{\partial \psi_l^m}{\partial t} &= A(\psi, \chi)_l^m - 2\Omega \left( \sin \phi \left( -\frac{l(l+1)}{a^2} \right) \chi_l^m + \frac{V_l^m}{a} \right) \\
-\frac{l(l+1)}{a^2} \frac{\partial \chi_l^m}{\partial t} &= B(\psi, \chi)_l^m + 2\Omega \left( \sin \phi \left( -\frac{l(l+1)}{a^2} \right) \psi_l^m - \frac{U_l^m}{a} \right) + E(\psi, \chi)_l^m - \left( -\frac{l(l+1)}{a^2} \right) (\Phi')_l^m \\
\frac{\partial (\Phi')_l^m}{\partial t} &= C(\psi, \chi)_l^m - \bar{\Phi} \left( -\frac{l(l+1)}{a^2} \right) \chi_l^m
\end{aligned}$$

Discretize in time by choosing a time scheme (for instance leap-frog)

$$\begin{aligned}
\psi_l^m(t + \Delta t) &= \psi_l^m(t - \Delta t) + 2\Delta t \left[ -\frac{a^2}{l(l+1)} A(\psi, \chi)_l^m - 2\Omega \left( \sin \phi \chi_l^m - \frac{a^2}{l(l+1)} \frac{V_l^m}{a} \right) \right] (t) \\
\chi_l^m(t + \Delta t) &= \chi_l^m(t - \Delta t) - 2\Delta t \left[ \frac{a^2}{l(l+1)} B(\psi, \chi)_l^m + 2\Omega \left( \sin \phi \psi_l^m + \frac{a^2}{l(l+1)} \frac{U_l^m}{a} \right) - \frac{a^2}{l(l+1)} E(\psi, \chi)_l^m - (\Phi')_l^m \right] (t) \\
(\Phi')_l^m(t + \Delta t) &= (\Phi')_l^m(t - \Delta t) + 2\Delta t \left[ C(\psi, \chi)_l^m - \bar{\Phi} \left( -\frac{l(l+1)}{a^2} \right) \chi_l^m \right] (t)
\end{aligned}$$

Figure 3.19: Treating nonlinearity with a time discretization with Leap-Frog scheme

The problem is to find a method to obtain the spectral coefficients of the nonlinear terms. In general the spectral coefficients can be obtained in a similar way to the Fourier coefficients,

projecting on a specific harmonic  $Y_k^j$  and integrating.

$$\int_{-\pi/2}^{\pi/2} \cos \phi d\phi \int_0^{2\pi} d\lambda \Psi(\phi, \lambda) Y_k^j(\phi, \lambda) = \int_{-\pi/2}^{\pi/2} \cos \phi d\phi \int_0^{2\pi} d\lambda \sum_{m=-J}^J \sum_{l=|m|}^{|m|+J} \Psi_l^m \delta_{kl} \delta_{mj} = \Psi_k^j$$

so the coefficients can be obtained by integrating the original (physical space) functions with the harmonics. The integration can be separated:

$$\Psi_k^j = \int_{-\pi/2}^{\pi/2} \cos \phi P_k^j(\sin \phi) d\phi \int_0^{2\pi} \Psi(\phi, \lambda) e^{ij\lambda} d\lambda \quad (3.130)$$

where the first integral is the *polynomial integration*. The second is for the *Fourier transform* with dependency on the latitude, so must be done at every latitude over longitude, the result of this Fourier integration is a vector of numbers at any latitude, this is used to make the longitudinal integration and you obtain the coefficients.

TABLE 1.—Transform model timing

Wave number truncation, $J$ (rhomboidal)	Transform grid	Number of Gaussian latitudes	Number of points per latitude circle	Computation time/time step (s)
7	20	24	24	2.3
12	32	48	48	7.8
15	40	48	48	12.2
24	64	96	96	39
30	80	96	96	67

TABLE 2.—Interaction coefficient timing

Wave number truncation, $J$ (rhomboidal)	Number of interaction coefficients*	Computation time/time step (s)
5	3,657	1.1
7	16,250	4.2
10	83,169	18.6
13	283,577	59
15	558,276	114

\*Comprising interaction coefficients arising from scalar triple products and from ordinary products.

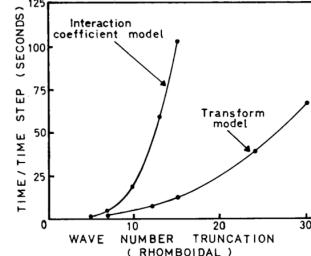


FIGURE 1.—Computation time per time step (s) as a function of spectral resolution. Integrations of a global spectral model employing a transform method and employing the interaction coefficient method are compared.

Figure 3.20: Gaussian grids

Numerically an integration is a summation, just like the derivatives are differences. However, we have freedom to choose the points and it can be shown that if the points are chosen in special points then the summation for polynomials is exact if enough points are selected. In the case of our models the integration is over latitudes, so we have to choose certain latitudes, these are known as *Gaussian Latitudes*. The minimum number of points is a function of truncation order.

A Gaussian grid is a type of non-uniform grid designed to minimize numerical errors when integrating over latitude bands on a spherical surface. The key feature of a Gaussian grid is that the grid points are more concentrated at the poles (where more detailed information is often needed) and become sparser toward the equator → commonly used in spectral methods to represent variables (such as temperature or wind) in terms of spherical harmonics. The Gaussian grid is particularly useful when integrating over the Earth's spherical surface.

## Numerical Transform

The discrete Fourier transform is computed numerically very efficiently by using a Fast Fourier transform (FFT). The integral over latitudes is a discrete Legendre transform and it requires the accurate discrete computation of the integral. It can be accomplished by a Gaussian quadrature at  $N$  special quadrature points  $\mu_i$ . Mathematically, the  $\mu_i$  are the roots of the ordinary Legendre polynomials, which are used to compute the  $N$ -latitudes of the Gaussian grid between the two poles

$$\int_{-1}^1 f(x) dx \approx \sum_i w_i f(\mu_i)$$

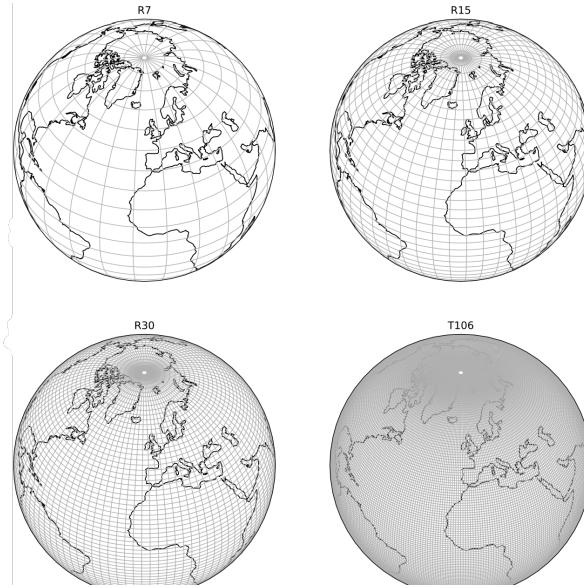


Figure 3.21: Examples of gaussian grids

Gauss showed that a  $N$ -point **Gaussian quadrature** yields an exact result for polynomials of degree  $2N - 1$  or less by a suitable choice of the nodes  $\mu_i$  and weights  $w_i$  for  $i = 1, \dots, N$ . Efficient algorithms have been developed to execute fast and exploit new architectures.

### 3.8.4 Truncation

The very first we have to do is choose how many function to use: the truncation order M and J. We have two indices, so we can arrange them in a plane.

Many climate models have used a rhomboidal truncation, R, of 15 waves numbers, i.e.,  $J=15$ , labeled R15. R15 gives a good approximation to the large-scale motions of the atmosphere but misses some of the smaller scale features. Increasing the resolution to R30 improves the representation to the original  $2.5^\circ$  data, while coarsening the resolution to R5 makes the pattern much too smooth.

It is not clear which type of spectral truncation is best. If triangular truncation (T) is used, then some high-frequency modes (small scales) are neglected. This truncation retains all modes where the total wave number is below a specified cutoff. If the resolution (number of grid points or spectral modes) is high enough, the type of truncation might not matter much because the neglected modes correspond to very small scales that have minimal influence on large-scale dynamics.

### 3.8.5 Summary

Spectral models are a class of numerical methods used in climate system modeling to represent atmospheric, oceanic, and other geophysical phenomena. They are particularly effective for solving partial differential equations (PDEs) that describe fluid dynamics (such as the equations governing atmospheric or oceanic flow) by using mathematical functions in the form of Fourier series or spherical harmonics, rather than on a grid of points as in traditional finite difference or finite element models.

Spectral models represent physical variables (e.g., temperature, pressure, wind speed) as a sum of basic functions, which are typically *Fourier series* (in the case of periodic domains) or *spherical harmonics* (in the case of global models on a spherical surface like the Earth). This

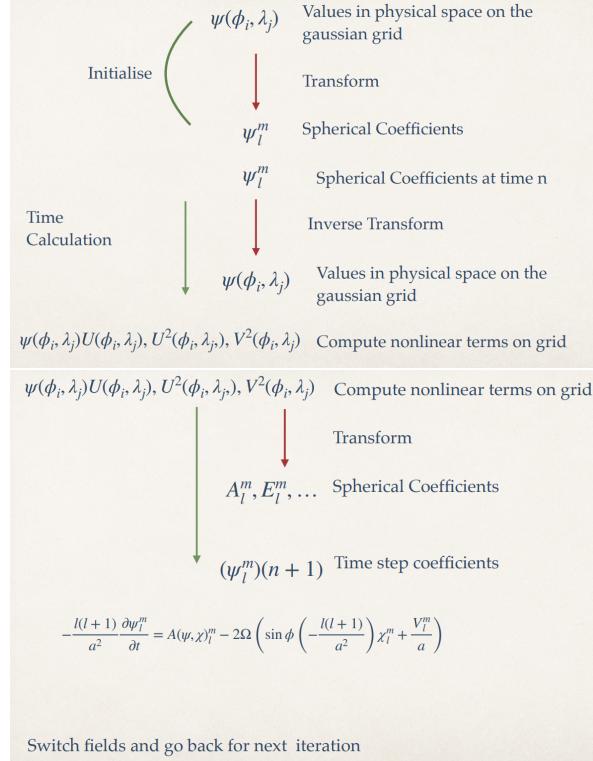


Figure 3.22: The transform cycle in the time-stepping

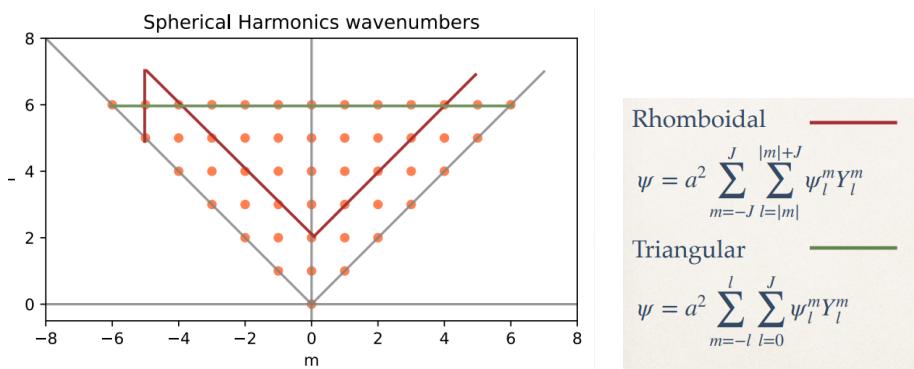
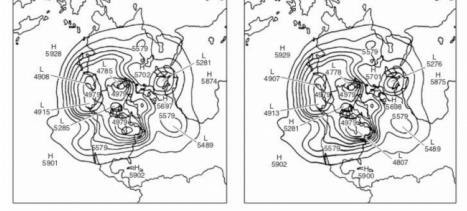
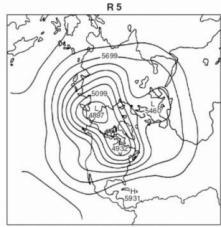
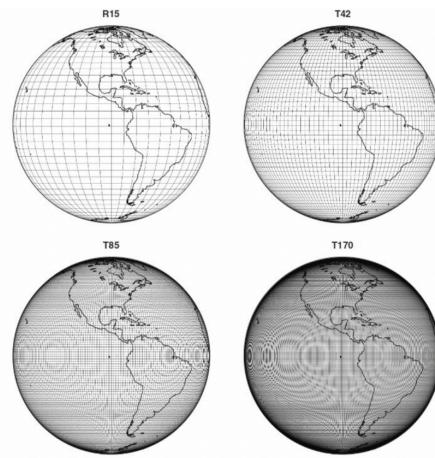


Figure 3.23: a) spherical harmonics wavenumbers; b) form of the truncations



**Figure 4.11** Demonstration of the effects of various horizontal truncations of 500 mb patterns of geopotential height ( $m$ ) of  $2.5^\circ$  latitude/longitude data: rhomboidal, R5, R15, and R30, and the original  $2.5^\circ$  data (All Waves). (Redrawn from David Baumhefner, personal communication, 1986.)



**Figure 4.12** Gaussian and triangular grids on the globe for various resolutions: rhomboidal, R15, and triangular, T42, T85 and T170. (David Williamson, personal communication, 2002.)

Figure 3.24: Effects of truncation

approach allows for highly accurate representations of large-scale patterns with relatively few computational elements.

### How to construct a spectral model

- **Governing Equations:** the physical equations that govern the climate system. For example, the governing equations for atmospheric dynamics are the *Navier-Stokes equations* or *primitive equations* (momentum, mass conservation, energy conservation).
- **Linearization:** in many spectral models, the equations are linearized or approximated to simplify the computational process (e.g.,  $\Phi = \bar{\Phi} + \Phi'$ ). This often involves assuming that deviations from a mean state are small and can be treated as perturbations.
- **Fourier or Spherical Harmonic Decomposition:** the physical fields are expanded in terms of a set of basis functions. For atmospheric and climate models on the Earth's surface, spherical harmonics are the most common. For a simpler system, Fourier transforms might be used for periodic domains (e.g., a wave equation on a flat domain).
- **Transformation to Spectral Space:** The variables (e.g., velocity components) are expressed in terms of their spectral coefficients in the Fourier or spherical harmonic basis. The system of equations then becomes a set of equations for these coefficients, which are typically simpler to solve than the original PDEs.

### Truncation of Spectral Models

The number of terms in the Fourier or spherical harmonic expansion is typically **finite**, and the expansion is **truncated** at a certain point to make the computation feasible. This truncation introduces a trade-off between *accuracy* and *computational cost*.

1. **Fourier Expansion:** in a Fourier expansion, the truncation involves limiting the number of sine and cosine terms that represent the spatial variability of a field. This results in a finite spectral resolution for the model.
2. **Spherical Harmonics Expansion:** for spherical models like global climate models, the variables (e.g., temperature, wind, pressure) are expanded in terms of **spherical harmonics**  $Y_l^m(\phi, \lambda)$ , which are functions defined on the surface of a sphere. These harmonics have two indices:

- (a)  $l$  is the degree,
- (b)  $m$  is known in this context as *East-West planetary wavenumber*, waves around a latitudinal circle (related to the longitudinal symmetry).

The expansion of a field  $f(\phi, \lambda)$  on the sphere is:

$$f(\phi, \lambda) = \sum_{l=0}^L \sum_{m=-l}^l a_{lm} Y_l^m(\phi, \lambda)$$

where  $L$  is the truncation level. The truncation level determines the maximum degree of spherical harmonics included in the model.

### Helmholtz theorem

You can decompose the velocity field into two scalar fields: the streamfunction  $\psi$  and the velocity potential  $\chi$ , this will simplify our analysis as a vector will transform as a vector changing its components, while a scalar will not change in different transformations. This theorem shows that you could represent a vector as function of two scalars (NB the gradient is a vector, see the Appendix??). The vector is depending on the meridional gradient of the streamfunction and vertical of the velocity potential.

The atmospheric is almost rotational, you can take the curl of  $V$  ( $\nabla \times V$ ) that's vorticity, described as the Laplacian of  $\chi$ . If the potential  $\chi$  is very small, then the atmosphere is dominated by the streamfunction  $\psi$ . This describes very well the atmospheric circulation outside of the tropics, there  $\chi$  becomes significant.

### Spherical Harmonics

Spherical harmonics are mathematical functions defined on the surface of a sphere. They are particularly useful in climate modeling because the Earth is roughly spherical, and they naturally represent the angular components of a field. More details can be found in ??.

- **Mathematical Form:** Spherical harmonics  $Y_l^m(\phi, \lambda)$  are solutions to the Laplace equation in spherical coordinates. They depend on two variables:

- $\phi$  (latitude, ranging from 0 to 180),
- $\lambda$  (longitude, ranging from 0 to 360).

The functions are orthogonal, meaning that integrals of different harmonic modes are zero unless they are the same mode (or its conjugate).

- **Properties:**

- The spherical harmonic functions form a complete, orthogonal set on the sphere, making them ideal for representing global fields like temperature, pressure, and wind patterns.
- The index  $l$  controls the **spatial scale** of the variations, with higher values of  $l$  corresponding to finer (smaller-scale) features,  $l \geq |m|$ .
- The index  $m$  corresponds to the **longitudinal symmetry** of the pattern. For a given  $l$ , there are  $2l + 1$  different  $m$ -values, ranging from  $-l$  to  $+l$ .

- **Truncation:** in spectral models, the expansion is truncated at a maximum value of  $J$  in our discussion. The number of coefficients increases rapidly as  $J$  increases because

the number of distinct  $m$ -values for each  $l$  is  $2l + 1$ . For example, the total number of coefficients for a given truncation level  $J$  is given by:

$$N_{\text{coeff}} = \sum_{l=0}^J (2l + 1)$$

which grows roughly as  $J^2$ . Truncating at a high  $J$  allows the model to capture finer details, but increases computational cost. EXAMPLE: For a truncation level  $J = 10$ , you would include all spherical harmonics up to degree 10. The number of coefficients would be:

$$N_{\text{coeff}} = \sum_{l=0}^{10} (2l + 1) = 1 + 3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 + 19 + 21 = 100$$

These coefficients are used to represent the field in terms of spherical harmonics.

### Advantages and Challenges of Spectral Methods

**Advantages** Accuracy with fewer points: spectral methods can represent large-scale features of the climate system (such as planetary waves, jet streams, etc.) very accurately with relatively few degrees of freedom (i.e., fewer spectral coefficients) compared to grid-based models → accuracy with fewer points. Also, for certain types of problems, spectral methods can be more efficient than finite difference methods because they avoid the need for fine grids over the entire domain → efficient computation.

### Challenges

- **Nonlinearity:** Spectral models can become more challenging when dealing with highly nonlinear phenomena (e.g., turbulence or convection). The nonlinear terms in the equations can introduce additional complexity, requiring special treatment like non-linear spectral operators.
- **Gridpoint-based variables:** While spectral methods work well for smooth solutions, they can struggle with sharp gradients or discontinuities, which are more naturally handled by grid-based methods.
- **Computational cost:** As the truncation level increases, the number of terms in the expansion grows quickly, increasing the computational cost. The choice of truncation level is often a balance between accuracy and available computational resources.

## 3.9 Lagrangian Methods

Remember the Gibbs phenomenon: the Fourier transform (Fourier series but with non-periodic functions with continuous range of frequencies) had problems in transforming functions with jumps. The Fourier transform of a function with a discontinuity produces a frequency spectrum that decays slowly, creating a high-frequency ringing in the reconstructed signal (when using an inverse Fourier transform). This ringing is analogous to the overshoot in Fourier series. It occurs because truncating the Fourier transform (e.g., ignoring very high frequencies) is equivalent to applying a low-pass filter, which introduces oscillations due to sharp edges in the signal.

This is a property not only of the Fourier transform but the all orthogonal transforms so the Legendre transform in the latitudinal direction will suffer from the same issue. This is a problem with any spectral model that will have difficulty in representing gradients steeper

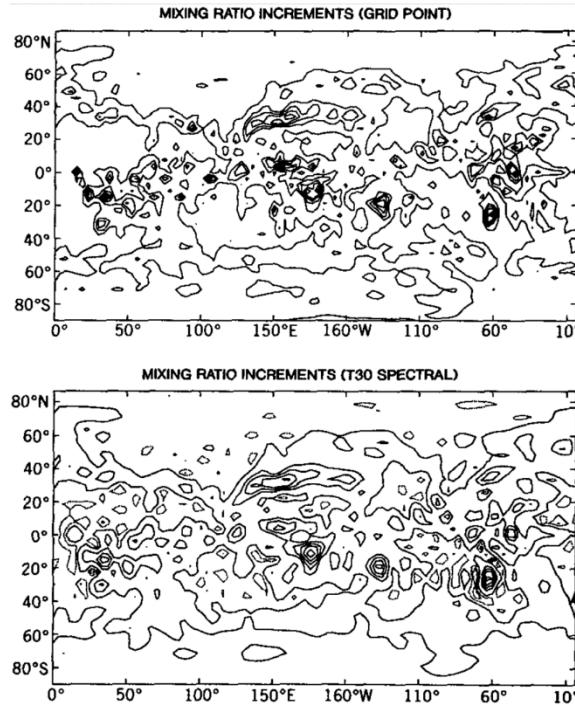


Figure 3.25: Lagrangian Methods

then the grid can sustain. Mountains will pose a problem, but also positive-definite quantities like water vapor and chemical tracers.

In an Eulerian advection scheme an observer watches the world evolve around her at a fixed geographical point. Such schemes work well on regular Cartesian meshes (facilitating vectorization and parallelization of the resulting code), but often lead to overly restrictive time steps due to considerations of computational stability. In a Lagrangian advection scheme an observer watches the world evolve around him as he travels with fluid particle. Such schemes can often use much larger time steps than Eulerian ones, but have the disadvantage that an initially regularly spaced set of particles will generally evolve to a highly irregularly spaced set at later times, and important features of the flow may consequently not be well represented. The idea behind semi-Lagrangian advection schemes is to try to get the best of both worlds: the regular resolution of Eulerian schemes and the enhanced stability of Lagrangian ones. This is achieved by using a different set of particles at each time step the set of particles being chosen such that they arrive exactly at the points of a regular Cartesian mesh at the end of the time step.

### 3.9.1 Semi-Lagrangian Methods

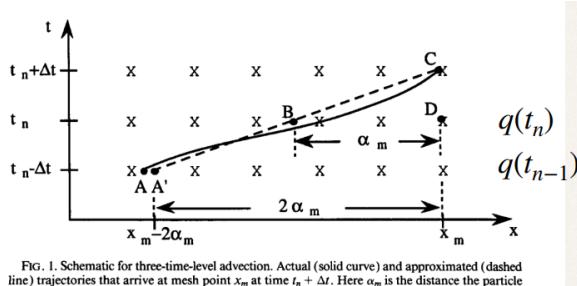


FIG. 1. Schematic for three-time-level advection. Actual (solid curve) and approximated (dashed line) trajectories that arrive at mesh point  $x_m$  at time  $t_n + \Delta t$ . Here  $\alpha_m$  is the distance the particle is displaced in  $x$  in time  $\Delta t$ .

Figure 3.26: Semi-Lagrangian method

Considering:

$$\frac{dq}{dt} = \frac{\partial q}{\partial t} + U \frac{\partial q}{\partial x} = S(x, t)$$

with  $\frac{dx}{dt} = U(x, t)$  Integrating along the trajectory  $A'C$ :

$$q(x_m, t_n + \Delta t) = q(x_m - 2\alpha_m, t_n - \Delta t) + 2\Delta t S(x - \alpha_m, t_n)$$

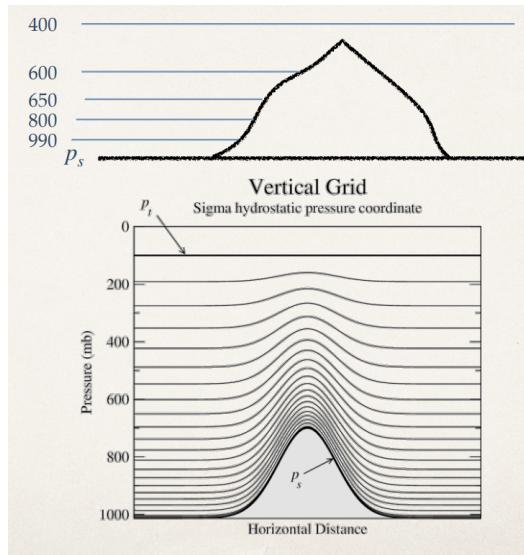
Assuming it is started at  $-2\alpha_m$ , and because it was transported by  $(t_n - \Delta t)$ , the  $S$  (source or sink term for  $q$ , where  $q$  will change along a trajectory only as a function of  $S$ ) is computed at  $t_n$  so at  $\alpha_m$  the displacement can be found solving:

$$\alpha_m = \Delta t U(x - \alpha_m, t_n)$$

and considering:

$$q(x - 2\alpha_m, t - \Delta t)$$

this quantity may not be at the grid point, I have to interpolate. This is preserving the trace and allowing me to look at a long time step. This transported quantities are used in a passive way because it doesn't reflect immediately in the rest of the dynamic. r0.58



## 3.10 Vertical coordinates

The original vertical coordinate was the Z coordinate that is simply the geometric height. However, as the atmosphere is in hydrostatic balance, one could use the pressure as a vertical coordinate.

$$(\lambda, \phi, z) \rightarrow (u, v, w)$$

$$(\lambda, \phi, p) \rightarrow (u, v, \dot{p} = \omega)$$

as

$$\nabla \cdot \mathbf{v} + \frac{\partial \omega}{\partial p} = 0$$

This implies to consider more layers in the vertical, each of these with different coefficients, then discretizing in the surface (in the making of models). However, spectral models require to have integrals over spheres: what happens if there are mountains? There will be boundary conditions. This choice hence as the problem that the surface pressure intersects the surface especially over mountains creating problems numerically in the definition of the boundary condition.

Phillips (1957) proposed to use a normalised pressure coordinate,

$$\sigma = \frac{p}{p_s}$$

The vertical velocity is now  $\dot{\sigma}$ . The spectral expansion might lead to bumps that propagate all the way up, they will create errors in the pressure gradient force. To have smoother bumps: another coordinate is introduced as in the figure 3.9.1. As the grid is staggered also in the vertical, in this way we'll get half layers: NB ground level is always an half-level.

**Hybrid coordinates** It is a sigma coordinates close to the ground, but smoothly transform into a pressure coordinate far from the ground

$$\sigma = \frac{p - p_T}{p_S - p_T}$$

# Chapter 4

## Climate Models

Last updated: 2024-12-13

Source file: [chapter-climate-models.tex](#)

We will use the shallow water equation, but to have a realistic case, we need to put it on the sphere. We start from the usual equation, for convenience we introduce the geopotential  $\Phi = gH$ :

$$\frac{\partial u}{\partial t} = -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - \frac{\partial \Phi}{\partial x} \quad (4.1)$$

$$\frac{\partial v}{\partial t} = -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - \frac{\partial \Phi}{\partial y} \quad (4.2)$$

$$\frac{\partial \Phi}{\partial t} = -\frac{\partial u \Phi}{\partial x} - \frac{\partial v \Phi}{\partial y} \quad (4.3)$$

Let's use  $U = u \cos \phi$ ,  $V = v \cos \phi$ ,  $\mathbf{V} = \hat{i}U + \hat{j}V$ . Until now all the methods we have introduced are designed to give us instruments to simulate and march the atmosphere and the oceans and, probably, much more. What is the difference between all these applications?

- Forecast (the target is the instantaneous state of the system) We start at a specific moment that we try to specify as accurately as we can and we want desperately to stay close to the state of the system for as long as we can. Here the main forcing is the internal dynamic of the atmosphere and the ocean. The conservation properties becomes progressively more important as the range of the forecasts becomes longer.
- Climate (the target is the statistics of the system) We want to reproduce its statistical distributions and the initial condition may not be so important, in fact we could start at rest.

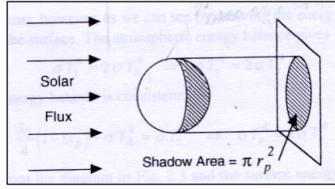
Here the main forcing are the external forcing and the structure (composition) of the atmosphere and the ocean. Conservation properties are important. The question here is: what will be the state of the atmosphere and the ocean in where the initial conditions won't be independent?

Both systems are affected by forcing, but in the forecast case there is no equilibrium.

### 4.1 Radiative Model

The Sun is the primary forcing and climate is the state in equilibrium with it. Firstly then, we have to take into consideration the fact that there must be preserved balance between absorbed solar radiation and emitted planetary radiation. To compute the absorbed radiation we have to take into consideration the Solar constant  $S_0$  which is the energy per unit of area and time that reaches a perpendicular surface placed at a distance of 1 astronomical unit. Therefore,

the energy incident on the planet is given by the product of the solar constant and the area that the planet exposes to the incident radiation perpendicularly to it (as shown in Figure 4.1). r0.18



It is called shadow area and it is approximately  $\pi R^2$  where  $R$  is the radius of the planet taken into consideration. Given that the absorbed radiation is equal to  $S_0 = \pi R^2$  This absorbed radiation during all day is distributed along the entire planet, so along the entire sphere  $4\pi R^2$ ; so that it becomes

$$S_0 = \frac{\pi R^2}{4\pi R^2} = \frac{S_0}{4}$$

Furthermore, it must be kept in mind that not all the incident radiation is absorbed; part of it will be reflected. The measure of a planet's reflectivity is called albedo. So that  $\frac{S_0}{4}(1 - \alpha)$  is the real absorbed radiation. The planetary albedo  $a$  is about 0.3, and the solar constant conventional value is 1361 W/m<sup>2</sup>

Solar Irradiance at the Planets			
Planet	Solar Irradiance, W*m <sup>-2</sup>		
	Mean	Perihelion	Aphelion
Mercury	9116.4	14447.5	6271.1
Venus	2611.0	2646.4	2575.7
Earth	1366.1	1412.5	1321.7
Mars	588.6	715.9	491.7
Jupiter	50.5	55.7	45.9
Saturn	15.04	16.76	13.53
Uranus	3.72	4.11	3.37
Neptune	1.510	1.515	1.507
Pluto	0.878	1.571	0.560

Figure 4.1: Solar Irradiance at the Planets

then

$$F_S = \frac{S(1 - a)}{4}$$

$$\sigma T^4 = \frac{S(1 - a)}{4} \quad \text{S-B Law}$$

$$T = \left( \frac{S(1 - a)}{4\sigma} \right)^{1/4}$$

with  $S = 1361$ ,  $a = 0.3$  and  $\sigma = 5.67 \cdot 10^{-8}$ , turns out that  $T_S = 254\text{K} = -18^\circ\text{C}$ . From Stefan Boltzmann Law,  $\sigma T^4$  is the emitted radiation by a blackbody at temperature  $T$ ; as  $T_S$  found this way is inaccurate, it can be concluded that the Earth-atmosphere system cannot be simply considered as a black body emitting at the temperature of the Earth's surface, since part of the emitted radiation is absorbed by the atmosphere itself. In this sense, the Earth's surface temperature does not reflect that of a black body in equilibrium with the incident solar radiation. However, the emission temperature obtained in this way is in good agreement with the temperature of the tropopause. The reason for this is that the tropopause represents the limit above which there is no significant absorption of infrared radiation, so the system, as seen from the tropopause, can indeed be approximated as a black body in terms of emission, in equilibrium with the incident solar radiation.

Satellite observations are giving us the observed long wave outgoing radiative flux from the planet outside the atmosphere. In fact it is so detailed that we have seen this flux geographically distributed over the all surface. Its global average is  $226 \text{ W/m}^2$  from ERA5. With the satellite observations, we can came up with an equation with the observed long wave outgoing radiative flux geographically distributed all over the surface.  $OLR_{mean}$  stands for Outgoing Longwave Radiation, the thermal radiation emitted by Earth back into space.  $OLR$  is a key component of the planetary energy balance, as it represents the primary way Earth loses heat to space. The amount of  $OLR$  depends on the Earth's surface temperature, atmospheric temperature profile, and the presence of greenhouse gases. We can use that flux to find the corresponding temperature:

$$\sigma T^4 = OLR_{mean} \implies T = \left( \frac{OLR}{\sigma} \right)^{1/4}$$

with  $OLR = 226$  and  $\sigma = 5.67 \cdot 10^{-8}$  hence,  $T_S = 251\text{K} = -21^\circ\text{C}$ . This is still not the temperature at the surface, whose global, long term mean is about  $288 \text{ K}$  or  $15^\circ \text{ C}$ , it is though what the satellite sees: it's the level at which emission reaches the out space. There is something that is making the surface opaque, hence the satellite doesn't see the surface but a layer on a higher level (remember that  $T$  decreases as the height increases). Maybe the earth surface is not really a black body and not all the radiation reaches outer space. Let us introduce a factor  $\tau$ , the transmissivity:

$$\tau \sigma T^4 = OLR_{mean}$$

and calibrate  $\tau$  with the observed surface temperature.

$$\tau = \left( \frac{OLR}{\sigma T_S^4} \right) \approx 0.58 \quad (4.4)$$

so the observed surface temperature can be explained only assuming that not all the radiation emitted from the surface reaches outer space. The level where the radiation is able to escape is above and it is cooler.

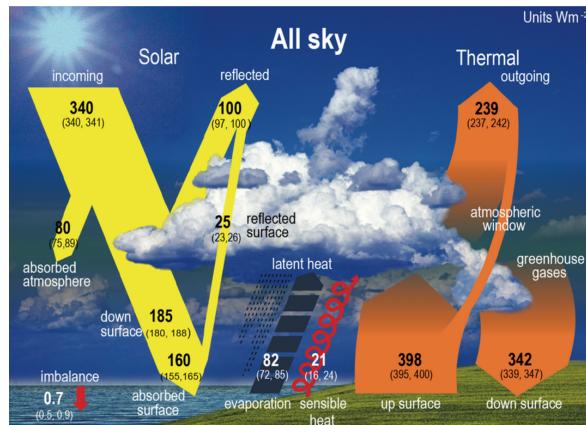


Figure 4.2: Earth-atmosphere system energy balance

The two- and three-layer atmosphere models are simplified frameworks to study Earth's radiation balance and the greenhouse effect. They explain how energy flows between the surface, the atmosphere, and space.

In the two-layer model, the Earth's surface absorbs solar radiation and emits infrared radiation, which is partially absorbed and re-emitted by greenhouse gases in the atmosphere. This re-emission back toward the surface enhances warming, maintaining an energy balance where incoming solar energy equals outgoing infrared radiation.

The three-layer model adds complexity by including an upper atmospheric layer (e.g., stratosphere), where radiation escapes to space. It considers energy transfer between the surface, middle atmosphere (troposphere), and upper atmosphere, capturing how each layer absorbs, emits, and redistributes radiation. The effective radiating level (ERL)—the altitude from which most radiation escapes to space—shifts higher with increased greenhouse gases, warming the surface while cooling upper layers.

These models illustrate key processes like vertical temperature profiles, with temperatures decreasing with altitude in the troposphere; radiative-convective equilibrium, balancing energy absorption, emission, and convection. They also depict climate sensitivity, showing how increased greenhouse gases amplify the greenhouse effect and raise surface temperatures. By simplifying Earth's energy flows, these models provide insights into how atmospheric layers interact and how greenhouse gases affect global warming.

#### 4.1.1 Two-layer atmosphere

We need both the in flux and the out flux to be at balance: assuming that the atmosphere is transparent so no change in  $S$  is caused by the atmosphere, we have the following equations for atmospheric balance:

$$(1 - \alpha)S = 4\sigma T_a^4 + 4(1 - \tau)\sigma T_S^4 \quad (4.5)$$

$\alpha$  is the albedo,  $\tau$  is transmissivity. The equation for the surface balance is:

$$\epsilon(1 - \alpha)S + 4\sigma T_a^4 = 4\sigma T_S^4 \quad (4.6)$$

where  $\epsilon$  is a factor accounting for possible absorption of the atmosphere at solar radiation,  $(1 - \alpha)S$  is carried from the atmosphere,  $4\sigma T_a^4$  is emitted by the atmosphere and  $4\sigma T_S^4$  emitted by the surface.  $T_a$  is the atmospheric temperature and  $T_S$  the surface temperature.

In the equation for the surface balance we account for two fluxes: one coming from the Sun, attenuated by a factor  $\epsilon$ , another coming from the atmosphere radiating downward (see the fig4.3). Hence,

$$\left( F_S = \frac{S\alpha\epsilon + S\alpha - S\epsilon - S}{4\tau - 8}, F_a = \frac{S\alpha\epsilon\tau - S\alpha\epsilon + S\alpha - S\epsilon\tau + S\epsilon - S}{4\tau - 8} \right) \quad (4.7)$$

this the solution in terms of the fluxes. Lets' look at different situations.

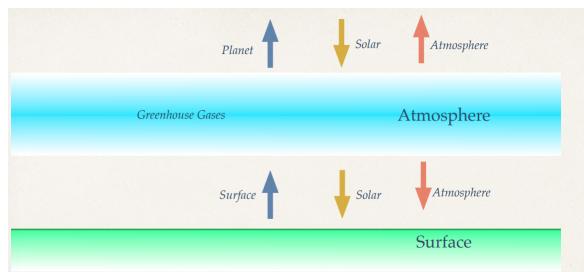


Figure 4.3: Two-layer atmosphere

- First assume the atmosphere does not absorb the solar radiation ( $\epsilon = 1$ ): the atmosphere is totally transparent to solar radiation; and the transmissivity of the atmosphere is 0.0, no emission from the atmosphere:

$$F_S = \frac{S(1 - \alpha)}{4} = 238 \text{ W/m}^2 \text{ and } F_a = 0$$

that is the same balance as before when we didn't account for atmosphere

2. Let's allow now for some absorption of terrestrial radiation and the transmissivity of the atmosphere is 0.5:

$$F_S = 396 \text{ W/m}^2 \quad F_a = 158 \text{ W/m}^2$$

$$T_S = 289 \text{ K} \approx 15^\circ\text{C} \quad T_a = 230 \text{ K} \approx -43^\circ\text{C}$$

meaning that even with less than 0.5, the presence of atmosphere will give  $T$  above 0°C. Opacity of the atmosphere is what makes possible the liquid water.

3. Assume a very opaque atmosphere the transmissivity is 1.0:

$$F_S = 476 \text{ W/m}^2 \quad F_a = 238 \text{ W/m}^2$$

$$T_S = 302 \text{ K} \approx 28^\circ\text{C} \quad T_a = 254 \text{ K} \approx -19^\circ\text{C}$$

**Opacity:** it refers to the degree to which the Earth's atmosphere prevents certain wavelengths of radiation (typically infrared radiation) from passing through it. This concept is closely connected to the greenhouse effect, as atmospheric opacity governs how much outgoing longwave radiation from the Earth's surface is absorbed and re-emitted by greenhouse gases. Greenhouse gases, such as carbon dioxide ( $CO_2$ ), water vapor ( $H_2O$ ), methane ( $CH_4$ ), and others, absorb infrared radiation emitted by the Earth's surface. The atmospheric opacity at certain infrared wavelengths increases with the concentration of these gases, preventing radiation from escaping to space. This absorption and re-emission process warms the Earth's surface and lower atmosphere.

Climate models use radiative transfer equations to simulate how variations in atmospheric opacity affect surface temperature, vertical temperature profiles, and energy transport.

#### 4.1.2 Three-layer atmosphere

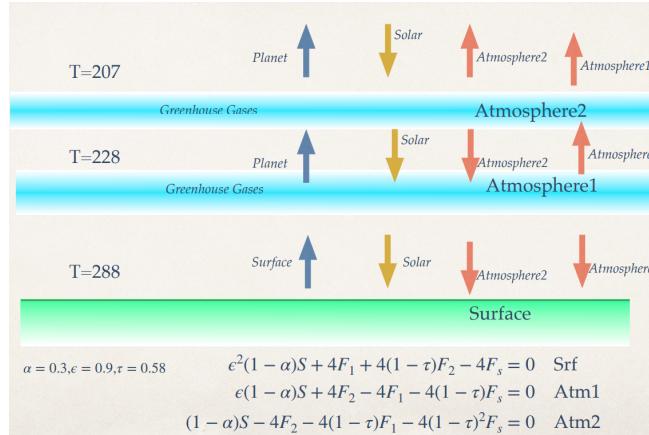


Figure 4.4: Three-level atmosphere flux balance

Remembering Kirchhoff's law (the emissivity and absorptivity of a real surface are equal for radiation with identical temperatures and wavelengths), we can introduce the temperature relations, we still consider the surface as a black body:

$$F_S = \sigma T^4 \quad F_1 = \tau \sigma T_1^4 \quad F_2 = \tau \sigma T_2^4$$

**Radiative equilibrium** The transfer equation describes the behavior in the continuous case:

$$\frac{dF^\downarrow}{d\tau} = B - F^\downarrow \quad \frac{dF^\uparrow}{d\tau} = F^\uparrow - B$$

where  $B = \sigma T^4$  is the blackbody emission,  $\tau$  the optical depth linked to the geometric height by  $d\tau = -e_L dz$ . If we are in continuous conditions the net flux is  $N = F^\uparrow - F^\downarrow$  and the long wave heating is proportional to the net flux divergence  $-\frac{\partial N}{\partial z}$ . So,

$$\frac{d}{d\tau} F^\downarrow e^\tau = Be^\tau \quad \frac{d}{d\tau} F^\uparrow e^{-\tau} = -Be^{-\tau}$$

At equilibrium there is no radiative heating, so:

$$\frac{\partial}{\partial z} (F^\uparrow - F^\downarrow) = 0 \rightarrow \frac{\partial}{\partial \tau} (F^\uparrow - F^\downarrow) = 0$$

in general this is not satisfied because the air is in motion, but suppose that such equilibrium exist, then the incoming solar radiation is balanced by OLR,  $F_T^\uparrow = F^\uparrow(\tau = 0) = S$ , so we get the boundary conditions at the top:

$$F^\downarrow = 0 \quad F^\uparrow = F_T^\uparrow \quad \text{at } \tau = 0$$

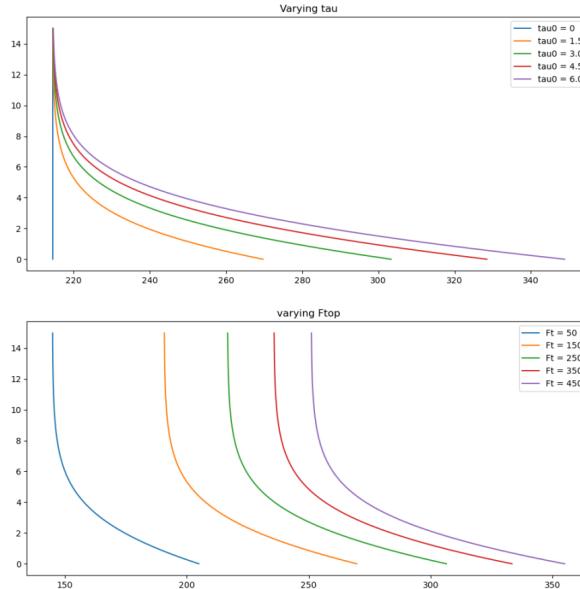


Figure 4.5: The change in OLR

then the solution is

$$F^\downarrow = \frac{\tau}{2} F_T^\uparrow \quad F^\uparrow = (1 + 1/\tau) F_T^\uparrow \quad B = \frac{(1 + \tau)}{2} F_T^\uparrow$$

The contribute from each level OLR is:

$$OLR_S = (1 - \tau)^2 \sigma T_S^4 \quad OLR_1 = \tau(1 - \tau) \sigma T_1^4 \quad OLR_2 = \tau \sigma T_2^4$$

add perturbation (change a little bit the opacity)

$$OLR_S = (1 - \tau - \Delta\tau)^2 \sigma T_S^4 \quad OLR_1 = (\tau + \Delta\tau)(1 - \tau - \Delta\tau) \sigma T_1^4 \quad OLR_2 = (\tau + \Delta\tau) \sigma T_2^4$$

we'll get different curves with different temperature.

Computing the change in *OLR*:

$$\Delta OLR_S \approx \Delta\tau[-2(1-\tau)]\sigma T_S^4 \quad \Delta OLR_1 \approx \Delta\tau(1-2\tau)\sigma T_1^4 \quad \Delta OLR_2 = \Delta\tau\sigma T_2^4$$

When greenhouse gas concentrations increase, the atmosphere becomes more opaque to infrared radiation, this reduces *OLR* escaping to space at the top of the atmosphere. The difference between the before and after state of *OLR* due to increased opacity is the *radiative forcing*:

$$R \approx -\Delta OLR \quad (4.8)$$

Radiative forcing describes how changes in atmospheric composition (opacity) affect the Earth's radiation budget by altering *OLR*. It quantifies the change in Earth's energy balance caused by a specific perturbation, such as an increase in greenhouse gases. The reduction of *OLR* due to higher greenhouse gas opacity leads to positive radiative forcing, indicating an energy surplus, initiating global warming and affecting the climate system through feedback mechanisms; while negative radiative forcing results in cooling.

Radiative forcing initiates changes in Earth's climate system, which are then amplified or damped by feedback mechanisms, such as the water vapor feedback (positive). The relationship between radiative forcing and global temperature change is encapsulated in the concept of climate sensitivity, often expressed as the temperature increase for a doubling of  $CO_2$  (see later on).

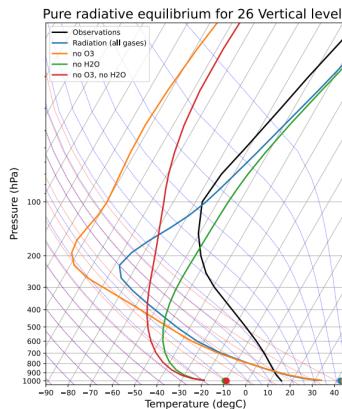


Figure 4.6: the black line corresponds to the observations; the blue to the component of the atmosphere, it decreases with  $T$  but in the stratosphere it changes inclination; the orange is without the ozone, responsible for the change of the slope in the stratosphere; the green one is without the water meaning that most of the opacity of the atmosphere disappears suddenly, at the bottom  $T$  is very cold; red without ozone and water, isotherm basically there is no atmosphere.

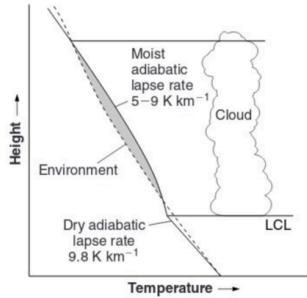
What about if we consider atmosphere composed by many more layers? In Figure 4.6 is shown a radiative equilibrium for 26 levels, with a realistic radiation calculation and observed components gases, with transmissivity and absorption coefficients ( $T$  profile with vertical coordinate  $p$ ).

## 4.2 Radiative-Convective Model

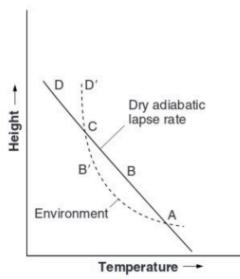
### 4.2.1 Radiative convective equilibrium

**Convection** Convection is about low layer mixing until they are gravitationally stable.

How to treat convection? Basically three steps:



**Figure 3.16** Schematic of cumulus cloud formation showing the lifting condensation level (LCL), dry and moist adiabatic lapse rates, and lapse rate for the air outside the cloud (labeled “environment”).



**Figure 3.17** Diagram of stable and unstable environmental conditions with respect to dry and moist adiabatic lapse rates. In each case the air between points  $A$  and  $C$  is unstable, while that outside the  $A - C$  segment is stable.

Figure 4.7: Schematic representation of convection

1. Identify the unstable layers;
2. Look at the temperature and moisture profiles;
3. Adjust until the profile is stable → conserve the energy distributing it vertically.

Radiative-convective equilibrium is a simplified framework for understanding the balance between radiative processes and convection in the atmosphere. The goal of this framework is to describe how the atmosphere redistributes energy vertically to achieve balance between the incoming solar radiation and *OLR*. In the troposphere for example convection plays a dominant role in redistributing heat, while radiation becomes the primary mechanism in the stratosphere.

To incorporate convection into a radiative model, we consider temperature and moisture profiles. Let  $T_r$ ,  $q_r$  be the equilibrium (reference) profiles for temperature and moisture. If the actual profiles  $T_i$  and  $q_i$  deviate from these references, the energy imbalance in the system can be calculated as:

$$\int_{p_B}^{p_T} c_p(T_r - T_i)dp \quad \text{temperature imbalance}$$

$$\int_{p_B}^{p_T} c_p(q_r - q_i)dp \quad \text{moisture imbalance}$$

where the integration is over the pressures from the bottom to the top of the convective layer. These expressions quantify the excess energy in the column relative to equilibrium, based on deviations in temperature and moisture.

If precipitation occurs, the total enthalpy (the sum of sensible and latent heat) in the system is conserved, and it is expressed as:

$$\int_{p_B}^{p_T} c_p(h_r - h_i)dp \quad h = c_pT + Lq$$

where  $L$  is the latent heat of vaporization. This conservation principle ensures that the redistribution of heat and moisture by convection does not alter the total energy content of the system. The vertical profiles are adjusted iteratively to stabilize the atmosphere. Before saturation, the temperature follows the dry adiabatic lapse rate, which describes how air cools as it expands. After saturation, the temperature follows the moist adiabatic lapse rate, which is slower because latent heat is released during condensation. Convective adjustment uses these principles to progressively bring the system closer to stability, ensuring that unstable layers are stabilized while conserving total enthalpy.

In the Radiative-Convective Equilibrium framework, convection and radiation interact to bring the atmosphere into equilibrium. This model provides a framework to understand how the atmosphere maintains energy balance. Convection redistributes energy vertically to stabilize the atmosphere, while radiation handles the energy exchange with space. The interplay between these processes is critical for maintaining the observed thermal structure of the atmosphere, particularly in the troposphere.

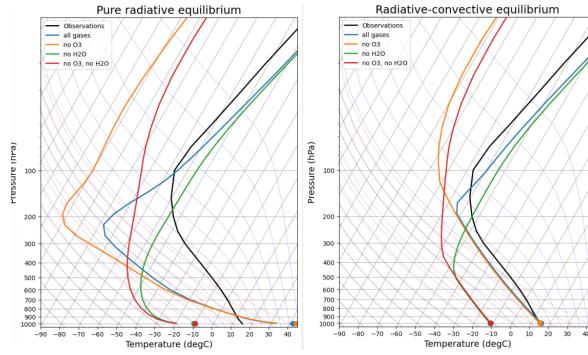


Figure 4.8: Radiative Convective equilibrium

### 4.3 Radiative forcing

Radiative forcing measures the change in energy balance in the Earth's climate system due to external factors (forcings). It quantifies the difference between the incoming solar radiation absorbed by the Earth and the outgoing infrared radiation emitted back to space.

Radiative forcing is a commonly used metric for assessing climate change mechanisms: compared to calculations of surface temperature change, it is a relatively straightforward calculation. It allows a more quantifiable and robust comparison of climate change mechanisms between models. The adjusted radiative forcing has been adopted by the Kyoto Protocol and IPCC (Intergovernmental Panel on Climate Change) process; they defined it as '*the change in net irradiance at the tropopause after allowing stratospheric temperatures to readjust to radiative equilibrium*'. The global-mean radiative forcing ( $\Delta R$ ) can be simply related to the equilibrium global-mean surface temperature change ( $\Delta T$ ) by the simple formula  $\Delta T = \lambda \Delta R$  which  $\lambda$  is the climate sensitivity parameter.

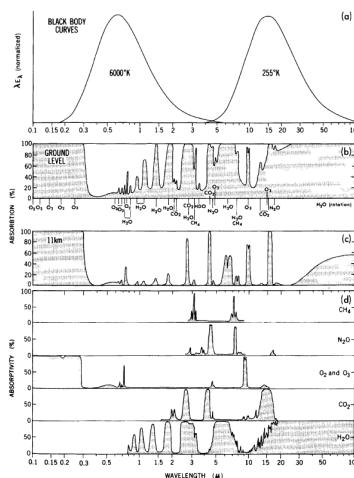
RF is defined also as the instantaneous perturbation in net radiative flux at the tropopause exerted by a change in a component of the radiative budget, with surface temperature and tropospheric state maintained in their unperturbed state. The term "radiative forcing" has been employed in the IPCC Assessments to denote an externally imposed perturbation in the radiative energy budget of the Earth's climate system. Such a perturbation can be brought about by secular changes in the concentrations of radiatively active species (e.g.,  $CO_2$ , aerosols), changes in the solar irradiance incident upon the planet, or other changes that affect the radiative energy absorbed by the surface (e.g., changes in surface reflection properties). This imbalance in the radiation budget has the potential to lead to changes in climate parameters and thus result in a new equilibrium state of the climate system. IPCC (1990, 1992, 1994) and the Second Assessment Report (IPCC, 1996) (hereafter SAR) used the following definition for the radiative forcing of the climate system: "**The radiative forcing of the surface-troposphere system due to the perturbation in or the introduction of an agent (say, a change in greenhouse gas concentrations) is the change in net (down minus up) irradiance (solar plus long-wave; in  $W/m^2$ ) at the tropopause AFTER allowing for stratospheric temperatures to readjust to radiative equilibrium, but with surface and tropospheric temperatures and state held fixed at**

the unperturbed values”.

The principal elements of the radiative forcing concept are summarised below:

1. In the one-dimensional radiative-convective model framework, the surface and troposphere are closely coupled, behaving as a single thermodynamic system under the joint control of radiative and convective processes, with a specified lapse rate determining the thermal structure. The stratospheric state is determined by the radiative equilibrium condition. The stratosphere and troposphere irradiances are together constrained by the requirement that the top of the atmosphere net total irradiance (i.e., radiative energy absorbed minus that emitted by the Earth’s entire climate system) must be zero at equilibrium. In applying the forcing concept to arbitrary spatial and seasonal time-scales, as opposed to the global annual mean, it has been assumed (WMO, 1992; SAR) that the stratosphere is in radiative-dynamical (rather than radiative) equilibrium.
2. The stratosphere is a *fast response system* which, in response to an imposed radiative perturbation, comes into equilibrium on a time-scale (about a few months) that is much more rapid than the surface-troposphere system (typically decades) (Hansen et al., 1997a; Shine and Forster, 1999). The latter is a *slow response system* owing principally to the thermal inertia of the oceans.
3. When a perturbation is applied (such as increases in well-mixed greenhouse gases), there is an *instantaneous* change in irradiances that manifests in general as a radiative imbalance (forcing) at the surface, tropopause and the top of the atmosphere. The rapid thermal re-equilibration of the stratosphere leads to an alteration of the radiative imbalance imposed on the surface- troposphere system (WMO, 1992), thereby yielding an *adjusted* forcing (SAR). The surface and troposphere, operating in a *slow response mode*, are still in a process of adjustment while the stratosphere has already reached its new equilibrium state. The SAR points out the clear distinction existing between the *instantaneous and adjusted* forcings.

R0.5



Let’s go back to the two layer model and compute the change in OLR:

$$R \approx -\Delta OLR_S - \Delta OLR_1 - \Delta OLR_2 = -\Delta\tau[-2(1-\tau)\sigma T_S^4 + (1-2\tau)\sigma T_1^4 + \sigma T_2^4] \quad (4.9)$$

Radiative forcing is the change in total radiative flux after a change in the absorbing properties of the atmosphere. Note that the radiative forcing depend on the vertical profile of temperature: for an isothermal profile \$T\_S = T\_1 = T\_2\$, hence:

$$R \approx 0$$

Let's try a case of radiative forcing due to an increase in greenhouse absorbers, using the tuned observed temperatures to the absorptivity parameter  $\tau$  that we calculate to be 0.58 for the observed atmosphere. For the tuned value of  $\tau = 0.58$ , we got that  $T_S = 288$  K,  $T_1 = 275$  K,  $T_2 = 230$  K let's compute the radiative forcing for an increase of 2% of  $\tau$ ,  $\Delta\tau \approx 0.012$ , then:

$$\Delta OLR_S \approx 3.78$$

$$\Delta OLR_1 \approx 0.65$$

$$\Delta OLR_2 \approx -1.86$$

So the total is 2.58 W/m<sup>2</sup>, so a small increase in absorption will decrease the outgoing *OLR*. Eventually the temperature at the ground will have to increase to adjust to the new situation. In short, radiative forcing is a direct measure of the amount that the Earth's energy budget is out of balance. For the Earth's climate system, it turns out that the level where this imbalance can most meaningfully be measured is the boundary between the troposphere and the stratosphere.

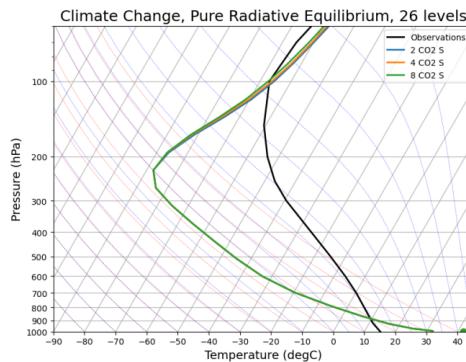


Figure 4.9: Use a radiative-convective model, but keep the troposphere at the unperturbed equilibrium values.

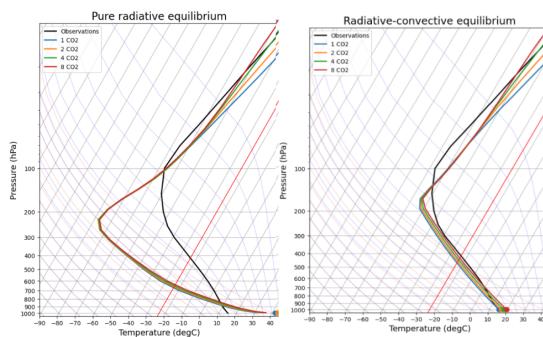


Figure 4.10: Climate change

### 4.3.1 Greenhouse feedback

It refers to processes in the climate system that amplify or dampen the warming caused by the greenhouse effect. The greenhouse effect is a natural process that warms the Earth's surface by trapping heat from the Sun. Solar radiation reaches the Earth, and the surface absorbs and emits heat as infrared radiation. Greenhouse gases in the atmosphere, such as carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>), water vapor (H<sub>2</sub>O), and others, absorb and re-radiate this infrared radiation, warming the Earth's surface and lower atmosphere. Without this effect, the Earth's average temperature would be much colder, making it uninhabitable.

While the natural greenhouse effect is essential for life, human activities have enhanced it, primarily through the burning of fossil fuels, deforestation, and industrial processes, which increase the concentration of greenhouse gases. This intensification of the greenhouse effect has led to global warming and climate change, causing rising temperatures, melting ice, sea level rise, and more frequent extreme weather events.

Positive feedback mechanisms increase climate sensitivity, potentially leading to faster and more severe warming; negative feedback mitigates sensitivity, stabilizing the climate system. The response parameter is:

$$\lambda_0 = \frac{\Delta R}{\Delta T_S} \quad \text{Wm}^{-2}\text{K}^{-1} \quad (4.10)$$

The change is essentially constant, but we have not taken into consideration the possible effect of water, that is an important component in creating the opacity. Water vapor creates powerful feedback in the climate system.

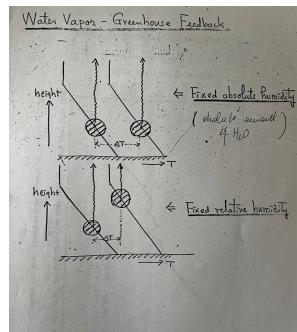


Figure 4.11: Greenhouse feedback from Manabe 1981

### 4.3.2 Water vapor feedback

Water vapor feedback is one of the most significant and well-understood feedback mechanisms in the Earth's climate system. It is a positive feedback that amplifies the warming caused by increased greenhouse gas concentrations. An initial radiative forcing (for instance an increase in  $CO_2$ ) generates a response in the surface temperature

$$\Delta T_0 = \frac{\Delta R}{\lambda_0} \quad (4.11)$$

When the climate system starts to adjust it will see a new radiative forcing:

$$\Delta R + f\lambda_0\Delta T_0$$

where  $f$  represents the fraction of the initial radiative forcing that is transferred to the next cycle, after  $n$  cycles we get

$$(1 + f + f^2 + f^3 + \dots) \lambda_0 \Delta T_0 = \lambda_0 \Delta T_0 \sum_n f^n$$

if  $f < 1$  the sum can be done and we get:

$$\Delta T = \Delta T_0 \frac{1}{1-f} = g \Delta T_0$$

so the overall effect is positive if  $f > 0$  and the system gain  $g$  is  $g > 1$ . This initial warming causes the surface temperature of the Earth to rise, leading to increased evaporation of water from oceans, lakes, and other sources. As more water enters the atmosphere in the form of vapor, the concentration of this potent greenhouse gas increases. Water vapor traps heat very

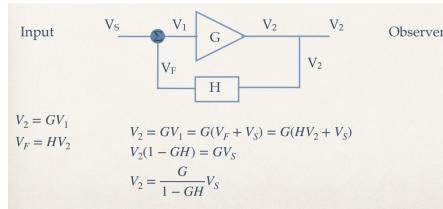


Figure 4.12: Feedback Loop, where  $G$  and  $H$  are machines that determine  $V_2$  that way.

effectively, which further warms the surface, creating a self-reinforcing cycle. The strength of water vapor feedback varies by region. It is most pronounced in the tropics, where the warm temperatures lead to significant evaporation and large increases in atmospheric water vapor. In contrast, in colder polar regions, the feedback is weaker because lower temperatures limit evaporation. This geographic variability means that water vapor feedback not only amplifies global warming but also contributes to differences in how warming manifests across the planet.

In nature, as in a complex GCM, water vapor tends to increase as the air temperature warms. The main reason for this is that the saturation-specific humidity (i.e. how much water vapor the air can hold) increases strongly with temperature. We can parameterize this effect in the column model by insisting that the relative humidity remain fixed as the column warms. In the radiative-convective model, the relative water vapor adjusts rapidly.

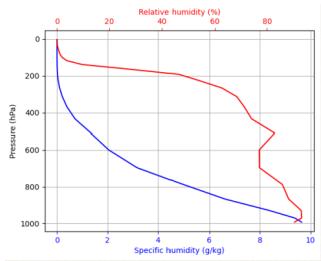


Figure 4.13: Relative water vapor in radiative-convective model.

The increasing temperature raises the saturation pressure allowing more water vapor to be present in the atmosphere. The previous perturbed experiment did not take into account this effect: it was basically a no-feedback simulation. To take into account the water-vapor feedback we have to constrain the model to keep its relative humidity fixed.

NB. Relative humidity is overall constant, specific humidity drops quickly with height and water vapor feedback increases the temperature of  $\sim 2^\circ\text{C}$ .

**Equilibrium Climate Sensitivity:** the temperature at the ground increases due to an increase of  $\text{CO}_2$ .  $T$  of the new experiment is 2.98 K, compare to the no-feedback experiment value of 1.31 K. The gain

$$g = \frac{ECS}{ECS_{\text{no feedback}}} > 2$$

Water vapor feedback doubles the warming of the ground, this was the major result by Manabe and Wetherald.

In his 1964 paper, ManabeManabe, 1964 focused on the interaction between radiation and convection in the atmosphere. The primary goal of the study was to understand how the atmosphere reaches a thermal equilibrium, considering both radiative and convective processes. Manabe developed a one-dimensional model that combined the radiative properties of the atmosphere with the effects of convective heat transfer. The model introduced the concept of

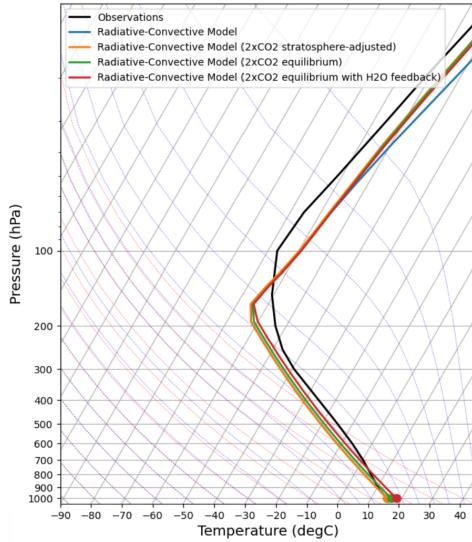


Figure 4.14: Model with water-vapor feedback: relative humidity fixed.

*convective adjustment*, a process where the atmosphere quickly adjusts to restore equilibrium when it is disturbed. Convective processes work to redistribute heat vertically, ensuring that the atmosphere can maintain thermal stability despite variations in surface temperatures and solar radiation.

Manabe demonstrated that the Earth's atmospheric temperature profile is governed by the balance between radiative cooling at higher altitudes and convective heating from the Earth's surface. The interaction between these two processes determines the temperature structure of the atmosphere, with convective adjustment helping to smooth out temperature gradients and reach a stable profile. Manabe also examined how changes in the concentration of greenhouse gases affect the temperature profile of the atmosphere. He showed that increased CO<sub>2</sub> leads to a warming of the lower atmosphere, as it enhances the greenhouse effect and reduces radiative cooling. One of the major conclusions of the study was that the atmosphere tends to reach a thermal equilibrium where the rates of incoming solar radiation and outgoing infrared radiation balance each other.

In his 1967 paper "Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity" [Manabe, 1967](#), Manabe extended his previous work on the thermal equilibrium of the atmosphere by incorporating the effects of relative humidity into his climate model. Assuming a specific distribution of relative humidity (the amount of moisture in the air relative to its capacity at a given temperature) allowed him to examine how water vapor, as a greenhouse gas, influences the thermal structure of the atmosphere and contributes to the energy balance. One of the most important results was the identification of the water vapor feedback mechanism. Manabe showed that as the temperature of the atmosphere increases, the amount of water vapor also increases, since warmer air can hold more moisture. This additional water vapor enhances the greenhouse effect, leading to further warming, which in turn increases the water vapor content. This feedback loop amplifies the initial warming caused by other factors such as increased CO<sub>2</sub>. The model demonstrated that the distribution of relative humidity significantly affects the vertical temperature profile of the atmosphere. Higher humidity in the lower atmosphere causes more water vapor to be present near the surface, which contributes to stronger greenhouse warming near the Earth's surface and a cooler stratosphere. The study showed how humidity influences both the radiative cooling and convective processes, ultimately shaping the thermal equilibrium of the atmosphere.

Building on his previous work, Manabe incorporated the concept of convective adjustment to account for the redistribution of heat and moisture in the atmosphere. The model showed that the presence of moisture, combined with the vertical heat transport due to convection, would lead to a different equilibrium state than in a dry atmosphere. This adjustment process, in turn, affects the thermal structure and stability of the atmosphere. Manabe also suggested that changes in the distribution of relative humidity, especially in response to global warming, could lead to nonlinear changes in the Earth's climate. For example, increased moisture in the atmosphere as a result of warming could accelerate the warming process, as water vapor is a powerful greenhouse gas. Initially general equations model for rain was produced when in a convective adjustment process, the specific humidity would be higher than the saturation of water vapor pressure, so the excess of humidity was considered rain. This was the first attempt to calculate in a physically consistent way the amount of rain. Rain is one of the major driving force for the atmosphere ( for the entire circulation).



# Chapter 5

## Ocean models

Last updated: 2024-12-13

Source file: [chapter-ocean-models.tex](#)

The modeling of the world's oceans for climate purposes is similar in many respects to the modeling of the atmosphere in terms of physical laws. However, there are key differences. The oceans involve slower processes and larger-scale movements, and their motions are influenced by the shape of the bottom topography (ocean floor). Radiative fluxes in the ocean are also more complicated due to how light interacts with water. For climate, the most important ocean dynamics are the ones that occur on large scales and over long periods of time, the ones responding to low-frequency, large-scale processes, emphasizing the importance of long-term, broad patterns over short-term variability. The time scales in the oceans are much longer than those in the atmosphere because of the greater thermal inertia; at the same time, the horizontal length scales are 1/10 of the atmosphere.

In the ocean, the hydrostatic approximation is an even better assumption. Furthermore, since sea water is nearly incompressible, the equation of continuity can be abbreviated by eliminating the time change of density, although density variations are generally included as far as buoyancy effects are concerned.

The Figure 5.1 shows the equation of the dynamics of the ocean.

Momentum conservation						
$\frac{\partial u}{\partial t} + \vec{u} \cdot \nabla u - fv = -\frac{1}{\rho_0} \frac{\partial P}{\partial x} + \nabla_h (K_{Mh} \nabla_h u) + \frac{\partial}{\partial z} \left( K_{Mv} \frac{\partial u}{\partial z} \right)$	$\frac{\partial v}{\partial t} + \vec{u} \cdot \nabla v + fu = -\frac{1}{\rho_0} \frac{\partial P}{\partial y} + \nabla_h (K_{Mh} \nabla_h v) + \frac{\partial}{\partial z} \left( K_{Mv} \frac{\partial v}{\partial z} \right)$	advection	Coriolis	Pressure gradient	Horizontal diffusion	Vertical diffusion
<hr/>						
<b>Hydrostatic</b>						
$0 = -\frac{\partial P}{\partial z} - \rho g$						
<b>Continuity</b>						
$0 = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}$						
<hr/>						
<b>Tracer conservation</b>						
$\frac{\partial T}{\partial t} + \vec{u} \cdot \nabla T = \nabla_h (K_{Th} \nabla_h T) + \frac{\partial}{\partial z} \left( K_{Tv} \frac{\partial T}{\partial z} \right)$						
$\frac{\partial S}{\partial t} + \vec{u} \cdot \nabla S = \nabla_h (K_{Sh} \nabla_h S) + \frac{\partial}{\partial z} \left( K_{Sv} \frac{\partial S}{\partial z} \right)$						
<hr/>						
<b>Equation of state</b>						
$\rho = \rho(S, T, p)$						

Figure 5.1: Equations in ocean dynamics

The equation of the state of the ocean is significantly different from the one of the atmosphere because it relates density to temperature, salinity, and pressure. The frictional term for the atmosphere is replaced by vertical and horizontal viscous terms. The hydrostatic equation is identical in form to that for the atmosphere. The first law of thermodynamics differs from that for the atmosphere because of the assumption of incompressibility, which allows the elimination of the term involving the time rate of change of density, and the replacement of the nonadiabatic term by explicit terms for the vertical and horizontal eddy diffusion of heat.

In addition, we have to take into consideration the fact that boundary conditions exist, like shown in Figure 5.2.

- surface: upper level of the oceans in direct contact with the atmosphere (water can't go up surface of the ocean magically)
- bottom: assuming that there is no flux that goes into the bottom of the ocean

<ul style="list-style-type: none"> <li>• Initial conditions</li> <li>• Surface boundary conditions (<math>z=\eta</math>):</li> </ul> $\frac{\partial \eta}{\partial t} = w \quad \text{Kinematic}$ $K_{Mv} \frac{\partial u}{\partial z} = \frac{\tau_x}{\rho_0} \quad \left. \begin{array}{l} \text{Wind stress} \\ \vdots \end{array} \right.$ $K_{Mv} \frac{\partial v}{\partial z} = \frac{\tau_y}{\rho_0} \quad \left. \begin{array}{l} \text{Wind stress} \\ \vdots \end{array} \right.$ $K_{Tv} \frac{\partial T}{\partial z} = \frac{Q}{\rho_0 C_p} \quad \text{Heat flux}$ $K_{Sv} \frac{\partial S}{\partial z} = \frac{S(E - P)}{\rho_0} \quad \left. \begin{array}{l} \text{Salt flux :} \\ \text{evap - rain} \end{array} \right.$	<ul style="list-style-type: none"> <li>• Bottom boundary conditions (<math>z=-H</math>):</li> </ul> $\vec{u} \cdot \nabla(-H) = w \quad \text{Kinematic}$ $K_{Mv} \frac{\partial u}{\partial z} = \frac{-C_d  \vec{u}  u}{\rho_0} \quad \left. \begin{array}{l} \text{Bottom} \\ \text{friction} \end{array} \right.$ $K_{Mv} \frac{\partial v}{\partial z} = \frac{-C_d  \vec{u}  v}{\rho_0} \quad \left. \begin{array}{l} \text{Bottom} \\ \text{friction} \end{array} \right.$ $K_{Tv} \frac{\partial T}{\partial z} = 0 \quad \text{Bottom-flux}$ $K_{Sv} \frac{\partial S}{\partial z} = 0 \quad \text{Bottom-flux}$
---	---

Figure 5.2: Boundary conditions

At the bottom of the ocean  $z = -H(\lambda, \phi)$ , here there is zero stress and no vertical flux of sensible heat or salinity. The boundary conditions, at the bottom on the vertical motion, are such that the vertical motions here, are induced by horizontal motions interacting with the bottom topography.

Fluxes of temperature are also described due to the fact that cooling is more effective in mixing the oceans and that has an impact on ocean dynamics. Fluxes of salinity (due to evaporation and precipitation) are indicated because an increase of it causes an increase of density.

There are different ways by which the atmosphere can force the ocean.

- transfer of momentum
- transfer of energy
- putting radiation fluxes on it
- evaporation - precipitation

For them, ocean response is related to the fact that if it is warm, it creates a pressure gradient and winds are created in response to it.

In 1969 Manabe and BryanManabe and Bryan, 1969 investigated the calculus obtained with a combined ocean-atmosphere model. Their paper presented a model combining oceanic and atmospheric components, a pioneering step in simulating their interactions to study climate change and variability. In their case, they are only allowed to use finite differencing and not spectral methods due to the boundary conditions. Some of the results are shown in Figure 5.3; here we can see the difference between the average calculated in the model and the one observed: there is an error of 5 ° that nowadays would be unacceptable.

It is easy to see that the temperature increases until mid-latitude and after that it decreases, stressing the important difference between the surface and the deeper part in homogeneity.

Global ocean models often describe their horizontal resolution with respect to their ability to permit or resolve mesoscale (i.e. Rossby radius scale) eddies.

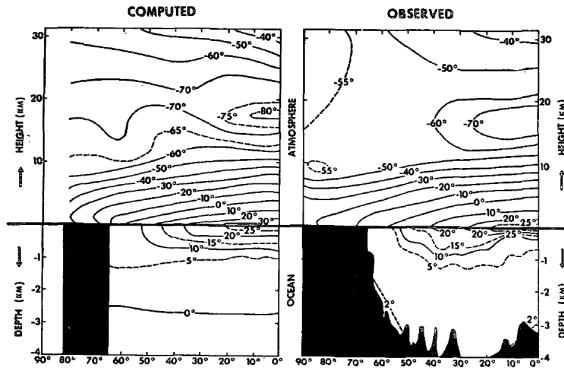


Figure 5.3: left-hand side: zonal mean temperature of the system represents the time mean over two-sevenths of the period of the final stage of the time integration; right-hand side: observed distribution in the Northern Hemisphere

**Problem with turbulence** Turbulence is a complex phenomenon, and modeling it accurately is critical for understanding ocean circulation. Turbulence in the ocean involves processes like vertical mixing, eddies, and small-scale currents, that cannot be resolved explicitly. Turbulence must be parametrized, meaning its effects are approximately based on larger-scale variables resolved in the model. The closure problem of the turbulence refers to the difficulty of determining how unresolved turbulent processes interact with larger-scale dynamics. The challenge lies in creating a parameterization scheme that effectively captures the impact of turbulence on heat, momentum, and salinity transport. In their work, Manabe and Bryan employed simplifications for turbulent mixing, focusing on large-scale processes like thermohaline circulation, while acknowledging that small-scale turbulence was not fully resolved in their model.

The Manabe and Bryan model laid the foundation for modern coupled climate models. The challenges they identified, including the closure problem for turbulence, remain central to ocean modeling.

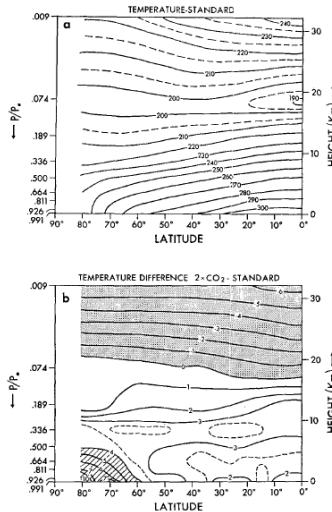


Figure 5.4: Manabe and Wetherald 1975

Manabe and Wetherald in 1975 [Manabe and Wetherald, 1975](#) tried to understand the effects of doubling  $CO_2$  concentration on the climate of a General Circulation Model. The results yield some indication of how the increase of  $CO_2$  concentration may affect the distribution of temperature in the atmosphere, in particular it raises in the troposphere and it

lowers in the stratosphere (Figure 5.4).

## 5.1 Ice models

They are somewhat different from ocean models because they don't consider a fluid anymore but a solid, discontinuous in nature, varying in its distribution from one time period to another. In contrast to ocean and atmospheric models, whose calculations center on determining the properties of the water/air, with sea ice the primary goal is to determine the location and each time step, ice thickness and aerial ice concentration.

The calculation for climate-oriented sea ice models can be divided into two categories:

1. calculations concerned with the thermodynamics of ice cover, the thickness and temperature structure of the ice, based on the principle of the conservation of energy.
2. calculations concerned with the ice dynamic: determine the motion of the ice based on the conservation of momentum.

(for more details on thermodynamics&dynamics pag 129-146 book)

## 5.2 The River Transport

It is an important aspect of the hydrologic cycle, although it is often not incorporated into climate models. In the real world, the fresh water that flows in the ocean as continental runoff affects the ocean salinity, and thereby the ocean density and current structure, as well as forming a vital link in the water cycle of the climate system.

The water balance model is usually part of the land vegetation surface component of the overall model and takes into account land cover type, monthly average of precipitation, potential evapotranspiration and temperature. Soil characteristics and topography are also inputs. Runoff has both surface and subsurface components. Eventually, the cumulative flow from the water transport model goes into the ocean.

## 5.3 Grids

Two possible grids: structured grids (cartesian, curvilinear) for global ocean, basin scale, coastal applications e.g. NEMO, MOM, ROMS, ecc; and unstructured grids, which domain is tiled using more general geometrical shapes (triangles, ...) pieced together to optimally fit details of the geometry → tidal modelling, near shores, estuaries e.g. FVCOM, FESOM, MPAS, ecc.

### 5.3.1 Regular grids

- regular spaced lines;
- on a spherical earth cannot have both uniform grid spacing and straight lines → grids tend to be curvilinear and the grid cell area tends to vary;
- lat-lon regular grids also have a problem at the poles where grid lines converge;
- they are computationally efficient;
- have a relatively straightforward analysis algorithm;
- have benefited from decades of research experience.

but for a given latitude they have a fixed resolution: to increase resolution near the edge of an ocean basin (where you want it!) requires an increase of resolution everywhere including out in the middle of the ocean (where you don't!). An example are the tripolar grids: regular grid laid over the high latitude North Hemisphere region with two rotated poles located over land.

### Unstructured grids

- irregular grids are designed to give more freedom to put spatial resolution where you want it
- a common scheme is composed as a series of triangles = finite elements
- by varying the size we can construct a non-uniform horizontal resolution over the computational domain
- they are efficient owing to the fact that resolution can be tailored to need as a function of space
- they can accurately represent highly irregular coastlines and topography (no need for "staircase" coasts and topography)

but they have unresolved issues with flows dominated by geostrophy and advection (characteristics of large-scale ocean flows), and they have spatially variable resolution-dependent physics (e.g. viscosity and diffusivity coefficients should really be resolution dependent).

### Adaptive Grids

- Unstructured meshes with a dynamically changing resolution responding to the nature of flow in time,
- they are very cool but very much in its infancy for ocean modeling applications.

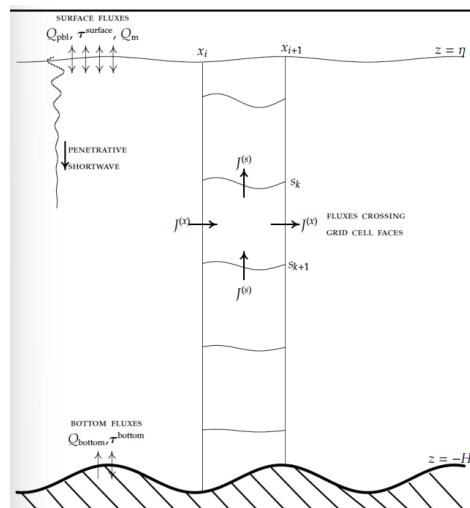


Figure 5.5: Vertical Grids

**Vertical Grids** The choice of the vertical coordinate system is loaded because:

- the oceans are forced at the surface (most of the action occurs there);
- they are strongly stratified and they are adiabatic in the interior;
- there is complex bottom bathymetry to deal with.

As a consequence there exist a number of approaches to choose from. A different choice for the vertical coordinates can be done:

1. Geopotential or Pressure: common for non-hydrostatic process modelling and large-scale climate modelling ([MITgcm](#), [MOM](#), [NEMO](#))
2. Isopycnal: clean representation of interior quasi-adiabatic flows and overflows ([GOLD](#), [HYCOM](#))
3. Sigma: common for shelf/coastal modelling ([ROMS](#))

**Geopotential Coordinates** Absolute depth/z-coordinate system. These are based on a series of depth levels, it is common to add vertical resolution near the surface by decreasing the spacing between the levels in the upper ocean relative to the deep. Most common method for global models.

- ADVANTAGES: simple to set up, computationally efficient, there are no pressure gradient errors.
- DISADVANTAGES: increased vertical resolution near the lateral boundaries (i.e. continental slopes) requires the addition of grid cells throughout the basin; spurious diapycnal mixing associated with the numerical advection scheme.

**Sigma** Terrain following or  $\sigma$ -coordinate system. It is based on the fractional depth, scaled from 0 to 1:  $0.01\sigma$  level is 1% of the depth of the ocean, while  $0.99\sigma$  level is at 99% of the depth of the ocean. Applications for shelves and coastal areas.

- ADVANTAGES: mimics the bathymetry and allows high resolution near the sea floor regardless of depth or proximity to land.
- DISADVANTAGES: pressure gradient errors, issues with spurious diapycnal mixing coming from the numerical advection scheme

**Isopycnal vertical coordinates** Density ( $\rho$ )/"layered models". The vertical grid is defined by density surfaces, it exploits the fact that below the mixed layer, ocean currents generally flow along surfaces of equal density (flow is "adiabatic").

- ADVANTAGES: simple, "exactly isopycnal"
- DISADVANTAGES: perform poorly where the ocean is less stratified (e.g. in shallow water); no resolution in an unstratified fluid; no mixed layer unless you tack one on; issues with entrainment.

**Hybrid coordinates** It seeks to optimize performance by combining the best suited system in different regions based on the dominant processes at work. HYCOM=z coordinates in the surface mixed layer and density coordinates in the stratified ocean below. Here the vertical coordinates evolve in both time and space as the depth of the mixed layer changes.

- ADVANTAGES: dynamically optimized coordinate system gives improved results.
- DISADVANTAGES: high computational costs.

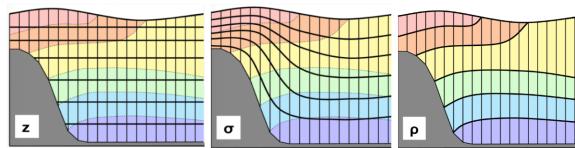
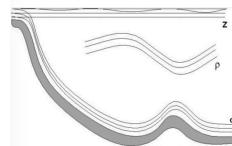
Figure 5.6:  $z$ ,  $\sigma$  and isopycnal coordinate systems

Figure 5.7: Hybrid coordinates

### 5.3.2 Resolution and mesoscale eddies

"Eddy-resolving" at  $1/10^\circ$  is not enough to resolve ocean "weather"

**Eddies** are swirling, circular movements of water in the ocean, resembling whirlpools, but they can vary greatly in size and duration. They are a fundamental aspect of ocean dynamics, transporting heat, nutrients, salt, and momentum across the ocean. Eddies often form due to shear instabilities (differences in velocity between layers), boundary currents (for instance Gulf Stream) and topography. The horizontal resolution needed to resolve the first baroclinic deformation radius with two grid points.

- There are different rules of thumb. One is that it takes 5 grid points to accurately define a feature without aliasing: this means  $1/8^\circ$  global resolution ( $\sim 15$  km) can accurately depict only features larger than  $\sim 60$  km;
- Models with variable grid spacing have variable resolution - beware of resolution dependent physics!
- Resolution is not cheap - because of the CFL<sup>1</sup> (NB. no transport faster than one grid cell per time step) condition, as we shrink the horizontal grid spacing we must add vertical layers and decrease the time step

At all (present-day) resolution, OGCMs resolve the mesoscale in some regions but not others.

Global ocean models often describe their horizontal resolution with respect to their ability to permit or resolve mesoscale (i.e. Rossby radius scale) eddies. Eddy-resolving does NOT mean all eddies are resolved or that all eddy effects of resolved eddies are acting. The spatial resolution of the ocean component of CMIP5/CMIP6<sup>2</sup> coupled models is  $0.2^\circ$  to  $2^\circ/0.1^\circ$  to  $1^\circ$ : from coarse (no eddies) to eddy-permitting (partially resolved eddy field). The effects of eddies need to be parameterized in coarse models (more later!); what to do in models that partially resolve the eddy field is an increasingly important question.

### Parametrizations

Processes need to be parametrized in a model for two reasons:

1. we don't spend the computational resources required to directly treat them because they are either too small or too complex.

<sup>1</sup>Courant–Friedrichs–Lewy stability condition

<sup>2</sup>see chapter 8

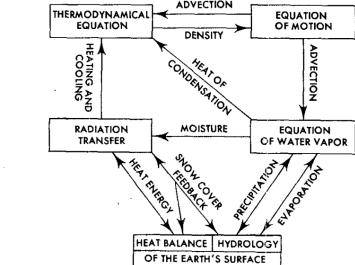
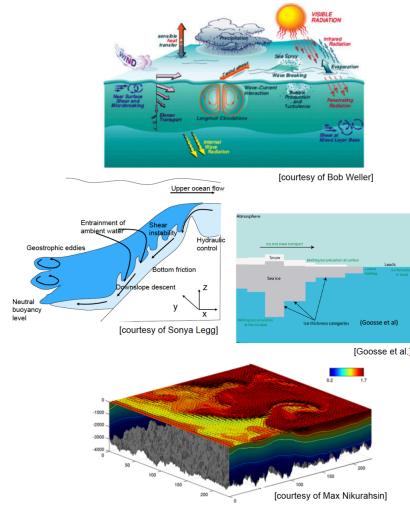


Fig. 2. Box diagram indicating the major components of the model.  
Arrows represent the links between components.

Figure 5.8: Fig.a: Parametrization and eddy problems. Fig.b: Major components of the model

2. we don't understand it well enough to be represented by an equation

Processes commonly parametrized in ocean models include: mesoscale eddy effects - submesoscale eddy effects - dense overflows - coastal processes - surface mixed layer processes - friction - sub-grid scale mixing - ocean-ice interactions.

*low resolution models:* the main problem is mesoscale eddies

*high resolution models:* submesoscale eddies (fronts and filaments), internal wave mixing, details of flowtopography interactions.

### Kind of eddies

The models must reproduce the eddy transport properties well otherwise that cannot maintain the basic flow.

**Blackmon 1976** M.L.Blackmon 1976 used spectral methods to decompose the variability of the 500 mbar geopotential height field into different spatial and temporal scales. He found:

- **Low-Frequency Variability (Planetary Waves):** Highlighted the dominance of quasi-stationary planetary-scale waves (large-scale undulations in the jet stream).
- **High-Frequency Variability (Synoptic Waves):** Identified transient synoptic-scale systems (e.g., midlatitude cyclones and anticyclones) as major contributors to the atmospheric variability.

He demonstrated that most of the variability occurs at intermediate (synoptic) scales. He quantified the distribution of energy across scales, helping to clarify the relative importance of large-scale versus small-scale atmospheric processes. He also found significant geographical differences in the distribution of wave activity, with distinct patterns in storm tracks over the Northern Hemisphere.

**Bjerknes** Jacob Bjerknes, the son of renowned meteorologist Vilhelm Bjerknes, is celebrated for his groundbreaking work in linking El Niño to the Southern Oscillation during the 1960s. While already recognized for his earlier contributions to the theoretical cyclone model developed with the "Bergen School" in the 1920s, Bjerknes turned his attention to understanding the dynamics of the equatorial Pacific.

Bjerknes observed that the sea surface temperatures (SSTs) in the eastern Pacific were unusually cold for such low latitudes, creating a sharp temperature contrast with the warm waters of the western Pacific. This temperature gradient drives a distinct atmospheric circulation, with cool, dry air moving westward along the surface and rising as warm, moist air in the western Pacific. Bjerknes named this systematic east-west atmospheric flow the Walker Circulation (see chapter 1 for more info).

He proposed that fluctuations in the Walker Circulation could trigger changes in the Southern Oscillation, setting off events associated with El Niño-Southern Oscillation (ENSO). Bjerknes described a feedback loop in which stronger easterly winds enhanced the upwelling of cold water in the east, intensifying the SST gradient and reinforcing the Walker Circulation. Conversely, weaker easterly winds reduced upwelling, diminished the SST gradient, and slowed the circulation. This mechanism explained the alternation between the warm El Niño phase, marked by a weakened Walker Circulation, and the normal cold state of the eastern Pacific associated with the high phase of the Southern Oscillation.

Bjerknes' insights established the foundation for modern understanding of ENSO, demonstrating how oceanic and atmospheric systems interact in a dynamic feedback loop to shape global climate patterns. The Walker Circulation is driven by the east-west SST gradient

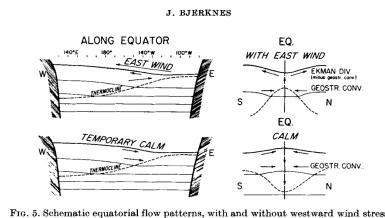


Figure 5.9: The Walker Circulation

along the equatorial Pacific: warm SSTs in the western Pacific generate low-pressure systems, promoting rising air and convection (clouds and rainfall); cool SSTs in the eastern Pacific lead to high-pressure systems, with descending air and dry conditions. This SST gradient establishes the pressure differences necessary for the Walker Circulation, with surface easterly winds flowing westward and upper-level winds returning eastward: stronger SST gradients reinforce the Walker Circulation by intensifying convection in the west and upwelling in the east. Weaker SST gradients, such as during El Niño events, disrupt the Walker Circulation by reducing pressure differences, weakening trade winds, and allowing warm water to spread eastward.

Climate forcing (e.g. global warming) modifies SST patterns, disrupting or enhancing the Walker Circulation. Changes in the Walker Circulation, in turn, influence SST distributions and climate change.

## 5.4 Caos

It is actually a misnomer, we should really talk about “sensitivity to initial conditions” (Tanner and Vattay, 2020). The idea had been fluctuated for some time both in Maxwell and Poincare’ work, probably the first colourful image was done by WS Franklin (1898). Ed Lorenz (1972) after seminal works in 1963 and 1969 introduced in 1972 at 139th Meeting of the AAAS, the concept ‘Does the flap of a butterfly’s wings in Brazil set off a tornado in Texas?’. Deterministic ideas dating to Newton and hyperdeterministic ideas by PS Laplace were put in discussion.

## Lorenz System

$$\frac{dx}{dt} = \sigma(y - x) \quad (5.1)$$

$$\frac{dy}{dt} = x(\rho - z) - y \quad (5.2)$$

$$\frac{dz}{dt} = xy - \beta z \quad (5.3)$$

The solutions of this system represent the traveling waves around a cycle of constant latitude.

$$\begin{aligned}\dot{x}_j &= x_{j-1}(x_{j+1} - x_{j-2}) - x_j + F \quad \text{for } j = 1, \dots, N \\ x_{j-N} &= x_{j+N} = x_j\end{aligned}$$

## 5.5 Vorticity dynamics

### 5.5.1 What maintains the circulation?

Starting from the basic structure of the wind, the jet is the factor maintaining the circulation. We need to write the equations of the jet stream. Consider the vorticity budget for a non-divergent case:

$$\frac{\partial \zeta}{\partial t} = -\mathbf{v} \cdot \nabla(\zeta + f) \quad (5.4)$$

and introduce now the zonal mean:

$$\bar{A} = \frac{1}{2\pi} \int_0^{2\pi} A d\lambda \quad (5.5)$$

separate each field in zonal mean and deviation,  $A = \bar{A} + A'$ :

$$\frac{\partial}{\partial t}(\bar{\zeta} + \zeta') = -(\bar{u} + u')\frac{\partial}{\partial x}(\bar{\zeta} + \zeta') - (\bar{v} + v')\frac{\partial}{\partial y}(\bar{\zeta} + \zeta' + f_0 + \beta y)$$

because it is tangent to the Earth and it needs linearization, zonally averaging again:

$$\frac{\partial \bar{\zeta}}{\partial t} = -\bar{u}\frac{\partial \bar{\zeta}}{\partial x} - \bar{v}\frac{\partial \bar{\zeta}}{\partial y} - \beta \bar{v} - \frac{\partial}{\partial y}\bar{v}'\zeta' \quad (5.6)$$

because  $\bar{A} = 0$ , subtracting this equation from the first we get an equation for the perturbation. Notice that because in the non-divergent case there exists a streamfunction, then  $v = \psi_x$  so  $\bar{v} = 0$  for the periodic boundary conditions:

$$\frac{\partial \zeta'}{\partial t} = -\bar{u}\frac{\partial \zeta'}{\partial x} - v'\frac{\partial \bar{\zeta}}{\partial y} - \beta v' - \frac{\partial}{\partial x}(u'\zeta') - \frac{\partial}{\partial y}(v'\zeta' - \bar{v}'\zeta')$$

for small perturbation we can linearize it:

$$\frac{\partial \zeta'}{\partial t} = -\bar{u}\frac{\partial \zeta'}{\partial x} - v'\frac{\partial \bar{\zeta}}{\partial y} - \beta v' \quad (5.7)$$

consider the mean equation 5.6,  $\bar{\zeta} = \bar{v}_x - \bar{u}_y$ , but the zonal mean of a total derivative in  $x$  is zero because of periodic boundary conditions, so  $\bar{\zeta} = -\bar{u}_y$ :

$$-\frac{\partial}{\partial y}\frac{\partial \bar{u}}{\partial t} = -\bar{u}\frac{\partial \bar{\zeta}}{\partial x} - \bar{v}\frac{\partial \bar{\zeta}}{\partial y} - \beta \bar{v} - \frac{\partial}{\partial y}\bar{v}'\zeta'$$

because it is not divergent, the mean meridional flow can be obtained from a streamfunction,  $v = \psi_x' \bar{v}$  is also zero, then

$$-\frac{\partial}{\partial y} \frac{\partial \bar{u}}{\partial t} = +\bar{u} \frac{\partial}{\partial x} \frac{\partial \bar{u}}{\partial y} - \frac{\partial}{\partial y} v' \zeta'$$

averaging again and remembering that the  $x$ -derivative of a zonally averages quantity is zero we get:

$$\begin{aligned}\frac{\partial \bar{u}}{\partial t} &= \bar{v}' \zeta' \\ \frac{\partial \bar{u}}{\partial t} &= -\frac{\partial}{\partial y} v' u'\end{aligned}$$

This relation shows that the acceleration of the mean flow is caused by the meridional fluxes of momentum. A similar relation can be found for the mean zonal temperature linked to the eddy heat flux.

The relation between eddies and jet streams holds also in the three-dimensional case in the quasi-geostrophic approximation. In this case however we have also non zero zonal meridional and vertical velocity ( $v, w$ ).

$$\frac{\partial \bar{u}}{\partial t} = f_0 \bar{v} - \frac{\partial}{\partial y} v' u'$$

we can then find similar relations for the zonal mean of the temperature:

$$\frac{\partial \bar{\theta}}{\partial t} = -N^2 \bar{w} - \frac{\partial}{\partial y} v' \theta' \quad (5.8)$$

The thermal wind relations link the variables:

$$f_0 \frac{\partial \bar{u}}{\partial z} = -\frac{\partial \bar{\theta}}{\partial y} \quad (5.9)$$

$$\frac{\partial \bar{v}}{\partial y} = -\frac{\partial \bar{w}}{\partial z} \quad (5.10)$$

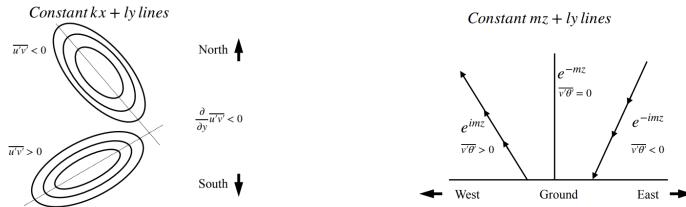


Figure 5.10: a) Momentum Flux. The meridional transport of momentum depends on the meridional slant of the eddies. b) Heat flux. The meridional transport of heat depends on the vertical slant of the eddies.



# Chapter 6

## Exploring Variability within Models

Last updated: 2024-12-13

Source file: [chapter-exploring-var.tex](#)

Horel and Wallace, 1981 **Horel1981**: their article was the first moment where people started looking at time series of monthly average values. People started to put together El Nino and southern oscillations: the role of the ocean was emerging. The EOF became the perfect instrument to do that: with the EOF you can decompose all the dataset on how they project (EOF gives the complete decomposition of the data), check [A](#) for more details:

$$X = U\Sigma V^T$$

where  $U$  is the matrix of the patterns,  $V$ , the matrix of the coefficients. When you plot the time coefficient of the first EOF, correlating the time series with the pressure all over the hemisphere, remote connections start to appear. The connections have different signs, like a wave was emerging from the pacific and propagating in the northern hemisphere. This prompt

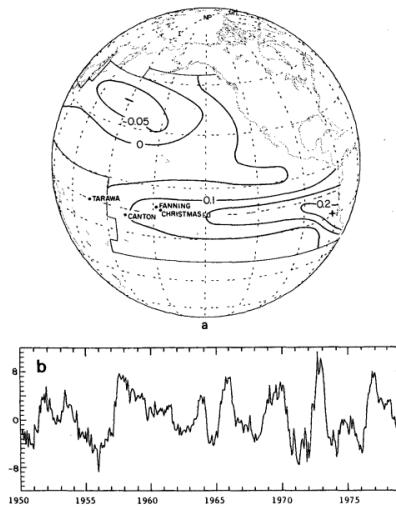


FIG. 1. (a) The first eigenvector of Pacific sea surface temperatures and (b) its corresponding time series, from Weare et al. (1976). For further details see text in Section 3. Also shown in (a) are the locations of the equatorial stations discussed in Section 4.

Figure 6.1: Horel and Wallace (1981)

was a systematic view with a month timescale variability: they used EOF to analyze 500 mbar data within a month's variability. There are 4 EOFs appearing during the analysis, the matrix had 45 columns (time series). As shown above [6.2](#), these patterns are a succession of positive and negative signs. In picture 16% of variance: reconstructing the timeserie using only that mode, the total variance of that mode will be 16% of the variance, or, having 45 modes, construct one by change, the change that it will look close to that one is 16%.

There was a network of connections disappearing beyond the synoptic view.

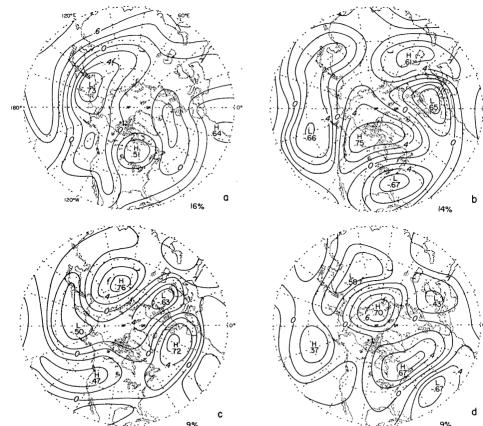


FIG. 27. Horizontal structure of the first four modes of an eigenvalue analysis expansion of 500 mb height as computed from the normalized covariance matrix  $\mathbf{R}$ . Based on the same 45 month data set as Fig. 7b. Numbers below each chart indicate the percentage of the hemispherically integrated normalized variance explained by that mode.

Figure 6.2: Correlation coefficients

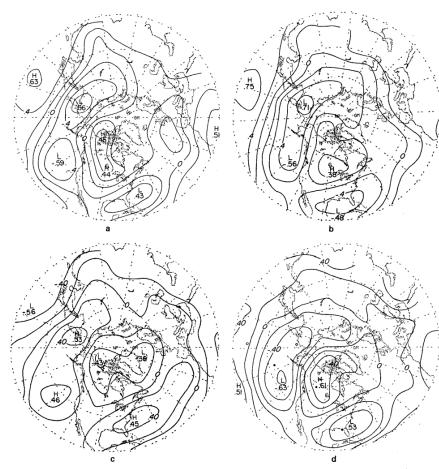


FIG. 9. Correlation coefficients between 700 mb geopotential height at gridpoints poleward of 20°N and (a) the Sea Surface Temperature Index, (b) December–February rainfall in Panama, (c) our Southern Oscillation Index, and (d) the tropical 200 mb index. Contour interval 0.2. The locations of the centers of action of the Pacific/North American and West Pacific patterns are denoted, respectively, by dots and open circles in Fig. 9a.

Figure 6.3: Connected variability in the ocean and in the atmosphere

Synoptic variation was understood as an instability of the basic flow. Nonlinear balance is reached if the instability grows very much outside the linear regime, and this could happen in a time series of a week/ten days. When you try to connect variability in the ocean with the variability in the atmosphere you get a pattern like the figure 6.3.

The idea came forward that at least a part of this teleconnection must be connected to the Ocean. There's a lot of variability in the ocean at monthly timescales. The figure below 6.4 shows deviations from seasonal timescales: a lot of variability at monthly timescales. The figure also shows the strongest El Nino ever recorded in 1993 and in 1998 another El Nino.

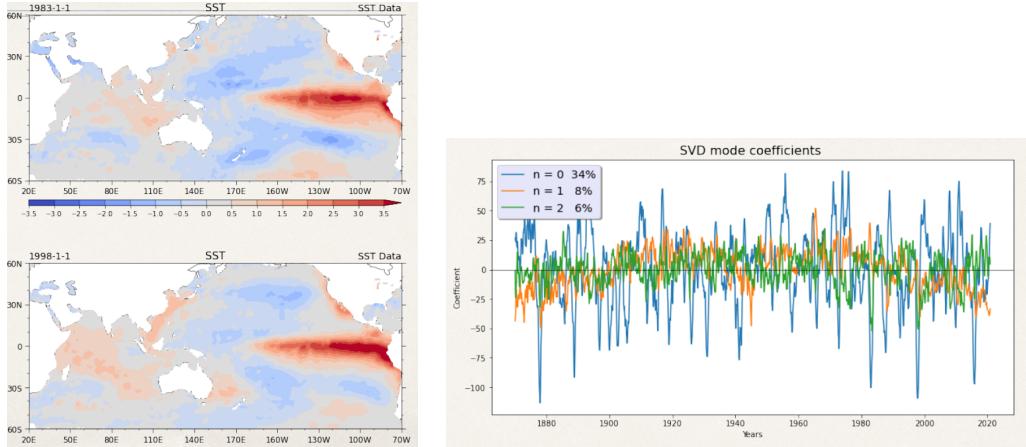


Figure 6.4: SST Variability over a monthly timescale and on the right main EOF timeseries

The right figure of 6.4 represents a modern analysis of SST variability from 1880 to last year. Once the notion that the SST was important became clear, people started to try to record the SST. However, there could have been a problem due to the difference in heat loss between different stations. In the figure,  $n = 0$  is the first EOF, here you can see huge oscillations (describes ENSO) and the negative phases of ENSO that are oscillating down. From now on started to come out with the idea that there is some periodic behavior.

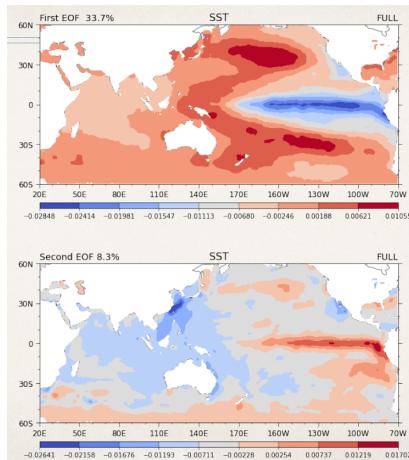


Figure 6.5: First two modes

In figure 6.5: the top one of El Nino, EOF cannot separate one phenomenon from the other: there are mathematical constraints (i.e. to be orthogonal to each other). In the following years, people try to make variations to address some of the problems. EOFs are very sensitive to where the variability is concentrated. As you go to higher modes, you get more structure for two main reasons: modes are getting more complicated and EOFs have to be orthogonal to each other because of their construction. Everything is dominated by the

large variations in the equatorial zone, and this variation in the North Pacific is relatively weighted less (here is very uniform but something is going on). Later it was discovered to be the Pacific decayed oscillation.

## 6.1 Designing a model for variability

All of these evidences were statistics, there is no reason to believe in causation. In order to demonstrate that it is the SST that forms those oscillations, we need to use a climate model and design an experiment. From evidence of past experiments it is clear that SST is connected to anomaly in the atmosphere, how should we design a model?

We want to design a numerical method, that is a simulation of climate using the model. For instance, we would like to see the evolution in time of SST and we have a dataset from 1880. Considering just the atmosphere, you need initial data in the model and boundary conditions in the ground (topography stressing the presence of mountains, it is not necessary to consider vegetation because it is part of the length scheme of the model, but we need a land-sea mask stressing where are lands, oceans and ice). You have to take into account the temperature of the ocean, in this case, SST (integrating forward in time will get a simulation in data). If you want to see the impact of the SST on the atmosphere you change the model inducing a disturbance (for example, changing the boundary conditions at the surface) and you get the monthly mean of SST. This type of experiment is called a *prescribed experiment*.

Run an atmospheric model using prescribed mean SSTs for the entire duration of the simulation. Within this framework, you can introduce an artificial perturbation to the SST in specific regions to observe its effects. For instance, the temperature at 2 meters above the surface (a standard atmospheric variable) is influenced by these SST changes, making it part of the atmospheric problem. By modifying the boundary conditions, such as increasing SSTs in certain areas, the model can reveal the resulting atmospheric responses to these changes.

So they built a strategy to do that: you start with a basic model and you add a perturbation to it. This will lead you to a comparison between one system with the perturbation and the other one without the perturbation, which is called the control system or the reference. Comparing the two systems, you'll gain the effect of the perturbation you applied.

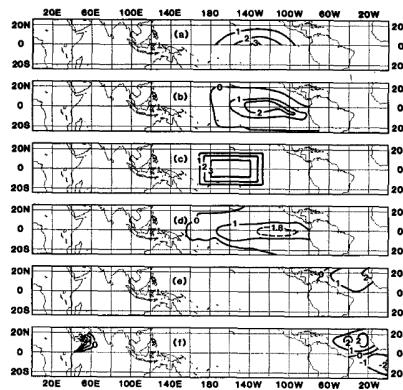


Figure 6.6: the fourth from the top is the paper

They did two perturbations and looked at the difference: the model would create a series of anomalies propagating. You create anticyclones in the north and the south of the anomaly. This fact provides an intense investigation of how the monthly scale can predict monthly anomalies. We are taking the difference between two strongly nonlinear simulations, so we are stretching the fact that we can still see the difference.

This means that the relation of the two simulations is quite linear so that the effect of perturbations can be identified. In detail, consider the geopotential at 300 mbar (high

atmosphere) 6.7, it is the streamfunction of the flow. The one on top is the control simulations,

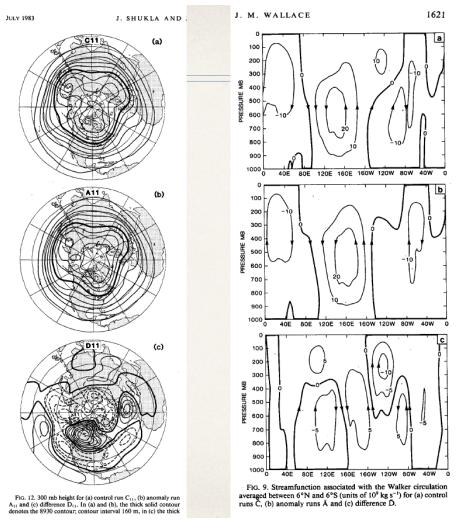


Figure 6.7: A section at the equator between 6N and 6S

the middle one is the perturbation experiment (with SST in the Pacific) and the bottom one is the difference between the two. Most of them are concentrated in the North Pacific, related to the SST anomaly in the equatorial Pacific. Note that when you look at the total field the differences are small but still exist. This was the first evidence that if you have perturbation that will have an impact, that creates a deviation from normal conditions.

Figure 6.7 is a section at the equator between 6N and 6S the arrows give the sense of circulation; in the difference, you see the extra rising air over the SST (central part or graph). This paper shows that the way SST is acting on the atmosphere is affecting the water circulation, especially increasing the vertical motion over the areas of increased SST. Why exactly air go up?

People decided to set up a special project called AMIP, an Atmospheric Model Intercomparison Project, based on the idea that SST is important but we have many models, expecting the SST in different ways, so how we could compare them? It was already clear that because of the sensitivity not only of the initial conditions but also to perturbation, it is not enough one simulation. We know that if we change a little bit initial conditions in any of these models, the evolution is slightly different, and also the response of SST it is; meaning that you have to do different experiments and confront them statistically.

They aren't using idealizing perturbation of the SST but the observed SST month by month. So it was designed a protocol in which all the people participating in the project, could run the models using the same dataset. As for the mean sea level pressure measured in

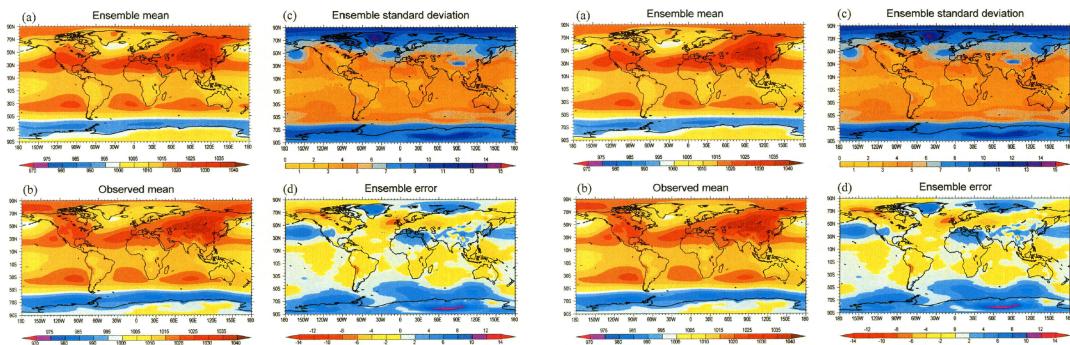


Figure 6.8: a)AMIP results for mean sea level pressure. b)AMIP result for precipitations.

winter between 1979 and 1988, despite the fact the ensemble means and the observed mean seem to be similar, there is an error of 14 mbar (good or bad depends on what you have to do). It was clear that there was a systematic problem: models seem to underestimate sea level pressure. If the error had a systematic structure, it could be possible that numerics weren't represented well. If we look at the standard deviation of the ensemble, we are looking at how these models would agree with each other (most of the deviation is presented in polar regions). In other words, we are looking at the degree of control that SST exerts on the atmosphere. If the control is perfect, all the models will give the same evolution; if not, models have the freedom to do slightly different things. The SST control is very effective in the tropical regions and in the mid-latitudes (east basin of the ocean). Weaker error occurs over North America, so SST seems capable of controlling the anomaly of circulation over it and the tropics; it seems not to have control over Europe (blob of high standard deviation).

Internal variation (for example, what will happen at 15 o'clock next month) beyond the range of initial conditions cannot be predicted.

The standard deviation is very small over the ocean, meaning the SST is controlling precipitations. In general, it's much more difficult than the dynamical field, because precipitations are much more difficult than other dynamics since they are driven by physical processes so you need to describe convection and turbulence very well. Other problems are linked with getting the observation and spatial correlation that is difficult. We get good data only recently, especially thanks to satellite like TRIMM that allow us to make estimates of precipitation also over the ocean.

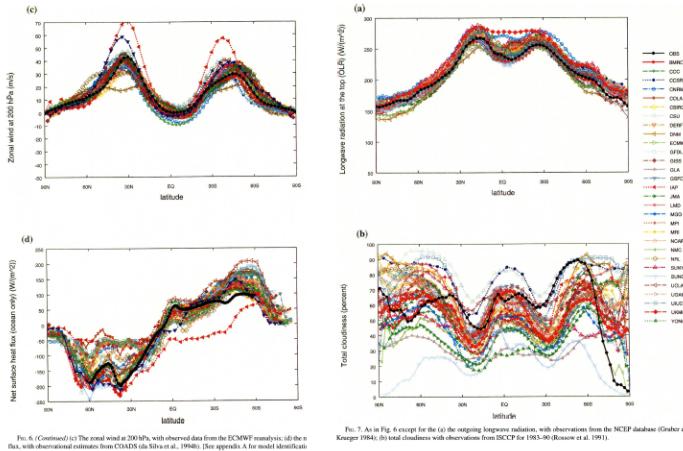


Figure 6.9: all the models participating in the comparison and the observations are the black line

All the models agree on the zonal wind, it means that the dynamics of zonal wind is easy. All the variations you see are how the model is formulated, you start to see differences because different formulation has a big impact, so how people choose to represent them has an impact on the result. If your balance at the top is not zero the model will cool or warm over 200-300 years

In the figure 6.9 the panel (b) is a disaster: clouds. That shows clouds are difficult to model.

This kind of result grew in people's determination to insert clouds in models because do not consider them to represent the issue of credibility.

FIG. 7. As in Fig. 6 except for the (a) outgoing longwave radiation, with observations from the NCEP database (Gruber and Krueger 1984); (b) total cloudiness with observations from ISCCP for 1985–90 (Ringer et al. 1991).

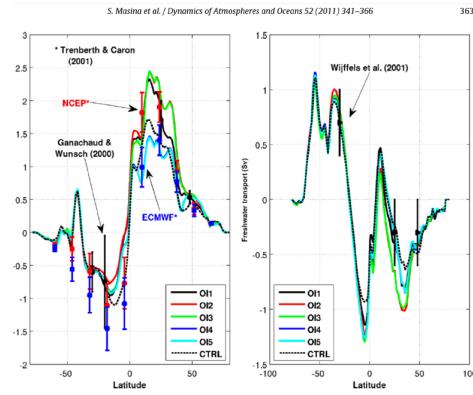


Figure 6.10: Climatological global heat transport and fresh-water transport: the different curves are different ocean re-analysis and black lines are observations

## 6.2 General Circulation Models

In the models, there are some variables called prognostic (expressed under time derivative, so they are in the left part of the equation), for atmospheric models, these are temperature, winds and water vapor; while for the ocean models, these are currents, temperature and salinity. Radiation instead, is computed from these variables so it is called a diagnostic variable. The only way to change temperature and salinity in the ocean is by the interaction between the atmosphere and the ocean.

Independently from which numerical methods, finite differencing or spectral, we use, we need grids. As we can see from the picture above 6.11, FAR had a resolution of 500 km

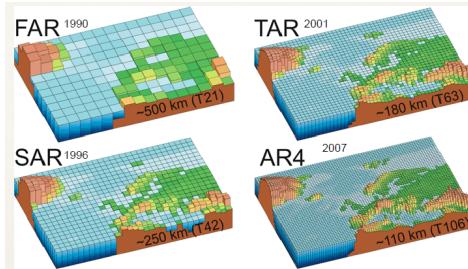


Figure 6.11: Evolution of different models

for finite differencing and T21 for the spectral method. Now instead we are able to reach a resolution of 30km x 30km and doing that made us capable of starting to detail the oceans.

It takes 5-6 years to collect all the info to have valid results to insert in the Assessment Report, a document that analyzes the state of the art of the science, collecting all the knowledge gained so far. If you want to do forecasting you want your model to follow exactly the observations. To do that, your initial conditions are important because for the atmospheric models they last 2-3 weeks, while for the oceans they last 12-50 weeks.

Some problems are derived from the observations since they do not cover all over the world uniformly and especially they are concentrated in the continents (which is the best density of station for the observations is still endless discussion, since they are expensive; probably the best thing you can do is to have maximum information with the minimum station distribution).

Really interesting is the ocean observing system which was possible to build thanks to the development in the knowledge of ENSO.

- TAO anchorage system to the sea floor in the equator measuring temperature

- ARGO FLOATS float on a systematic depth and every 10 days they change their buoyancy and rise up recording the profile of different variables, after that they go down again
- tidal station across the coast of the continents to measure sea level

Finally, with all this data collected, you have to fill them in both space and time, creating a physically consistent field of observations around the globe. Since everything you know is in the equations, you will use the models to do the simulations with different techniques developed through the years.

One important thing to take into consideration is that you have to take into account the ocean transport. It plays an important role in the climate system since it restores the imbalance of heat between tropics and polar regions. In particular, there is an excess of radiation energy in the equatorial region and a deficit in the polar region. In the picture below 6.12

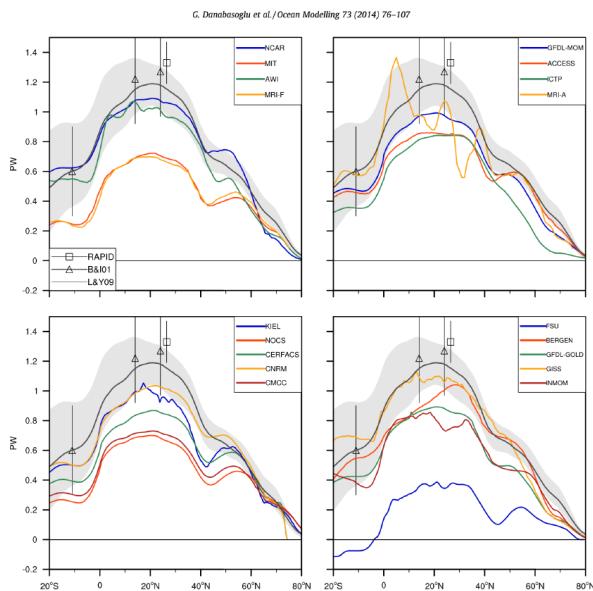


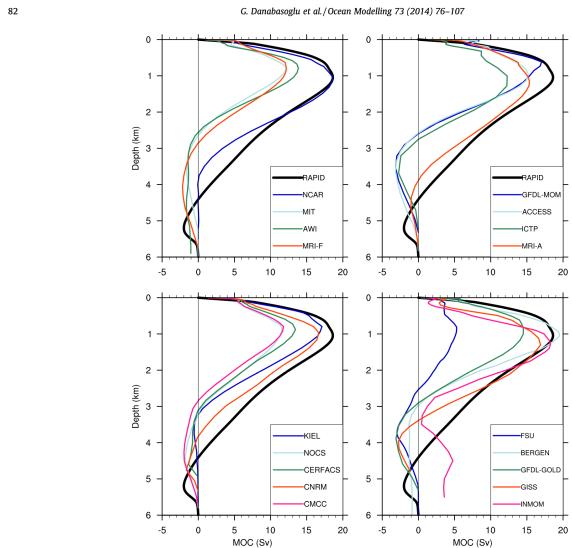
Figure 6.12: Time mean meridional heat transports for the Atlantic Ocean with different models

the observations are the point and all the lines are model simulations. All the models are underestimating transports. This graph shows us the main important thing in data assimilation: you have to reduce as better as possible the distance between observations and models.

Another important aspect that must be taken into consideration is the AMOC, Atlantic meridional overturning circulation. It is a part of the global circulation of the ocean that starts with the deep water in the North Atlantic where cold waters sink towards the bottom (due to its high salinity that promotes high density). Controlling the AMOC helps scientists to understand if this process really happens and how effective it is. If you stop the production of dense water, for example by freshening the North Atlantic with an increase of precipitation, all the AMOC will become weaker.

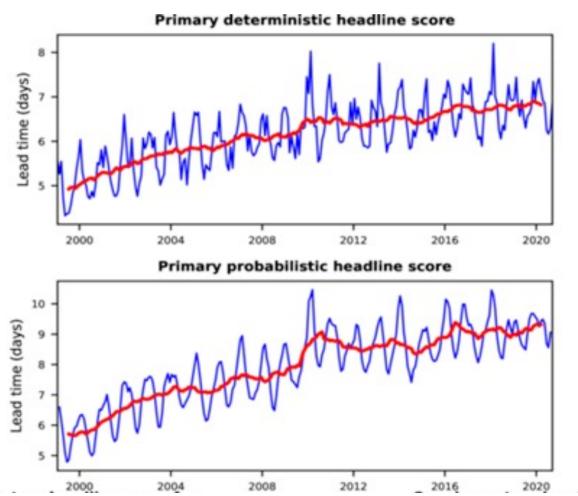
In recent years, scientists have developed new modes of investigation concerning simulations, experiments, and theories together. Also, improvement of forecasting skills helped them to improve the quality of simulations, specifically in the number of days in which the forecast stopped being useful.

Last but not least, different models have similar patterns of errors. This is a problem because, even if the mean of a variable is correct, I had to consider realistic variability from



**Fig. 5.** Years 2004–2007 mean AMOC depth profiles at 26.5°N from model solutions in comparison with the 4-year mean (April 2004–March 2008) RAPID data (thick black lines plotted in each panel).

Figure 6.13: NCAR doesn't really stress this phenomenon described above, despite the fact it follows observations quite well



**Figure 6.14:** Improvement in forecasting skills

it. If I improve the initial conditions, I'll improve the short timescale forecast; while if I improve the model, I'll improve the medium timescale forecast.

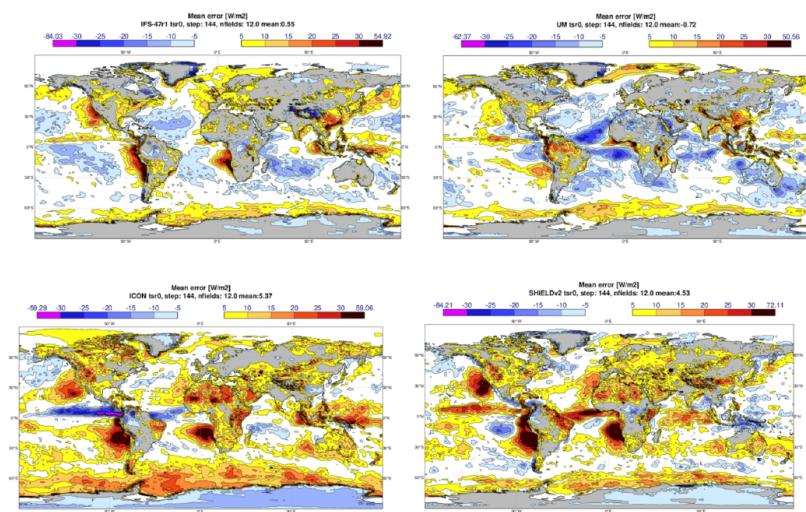


Figure 6.15: Example of mean errors in comparison different models

# Chapter 7

## Forcings

Last updated: 2024-12-13

Source file: [chapter-sst-forcing.tex](#)

### 7.1 SST forcing

Understanding SST variability and its role in shaping atmospheric anomalies is essential for seasonal climate forecasting. For example:

- Accurate SST forecasts for the next six months could provide critical insights into the likelihood of large-scale phenomena like El Niño, monsoons, or droughts, which are influenced by SST variability.
- Coupled ocean-atmosphere models that incorporate both internal dynamics and external SST-driven responses are critical for advancing our ability to predict climate anomalies over seasonal to interannual timescales.

#### 7.1.1 Climate forcings on SST

Climate forcing refers to external factors that influence the Earth's climate system, such as greenhouse gas emissions, volcanic eruptions, solar variability, and aerosols. These forces drive changes in sea surface temperatures (SSTs), which play a vital role in regulating global climate by influencing atmospheric circulation, ocean currents, and weather patterns. The interaction between climate forcing and SSTs involves complex mechanisms that include both direct effects and feedback processes.

One of the primary ways climate forcing impacts SSTs is through changes in radiative energy balance. For instance, increasing greenhouse gas concentrations traps outgoing long-wave radiation, causing the Earth's surface, including the oceans, to warm. Oceans absorb more than 90% of this excess heat, resulting in widespread SST increases. The effects are not uniform, as ocean dynamics influence regional variations in warming. For example, the Western Pacific Warm Pool retains more heat due to strong stratification, while upwelling regions, such as the eastern Pacific, may experience slower warming due to the cooling effects of vertical mixing and horizontal ocean currents.

Other forms of climate forcing also influence SSTs. Volcanic aerosols and human-emitted sulfate aerosols, for example, reflect incoming solar radiation, leading to localized or even global cooling of SSTs. The cooling effects are often more pronounced in the Northern Hemisphere, where industrial activity has historically been concentrated. In contrast, variations in solar radiation cause more subtle, global SST changes due to the ocean's high heat capacity.

**SST Changes as feedback to Climate Forcing** SST changes resulting from climate forcing do not only respond passively but also act as drivers of feedback mechanisms that amplify or modulate the effects of climate forcing. A key example is the water vapor feedback. Warmer SSTs increase evaporation rates, introducing more water vapor into the atmosphere. Since water vapor is a potent greenhouse gas, this enhances warming further. Similarly, SST changes influence cloud formation, which can either reflect sunlight back to space (cooling the Earth) or trap outgoing heat (warming the Earth). The balance of these cloud feedbacks is a significant area of ongoing research.

SST changes driven by climate forcing also amplify or modify major climate phenomena. One well-known example is the El Niño-Southern Oscillation (ENSO). Global warming is thought to alter SST patterns in the tropical Pacific, potentially affecting the frequency or intensity of El Niño and La Niña events. For instance, by flattening the SST gradient between the eastern and western Pacific, warming may weaken the Walker Circulation and favor more frequent El Niño-like conditions. Such shifts have significant implications for global rainfall patterns, droughts, and extreme weather events.

Changes in SSTs also influence large-scale ocean circulation patterns, such as the Atlantic Meridional Overturning Circulation (AMOC). Climate forcing that warms SSTs in the North Atlantic can disrupt this critical current system, leading to cascading effects on regional climates, including shifts in monsoons and weather systems across Europe and North America.

Long-term observations provide clear evidence of the impact of climate forcing on SSTs. Since the late 19th century, global SSTs have risen by approximately 0.8°C, primarily driven by greenhouse gas emissions. However, the warming is not evenly distributed. For example, the Arctic Ocean has experienced significantly faster warming due to the ice-albedo feedback, where melting ice exposes darker ocean surfaces that absorb more solar energy. Whereas, regions with persistent upwelling, such as the coast of Peru, have shown slower warming because cooler waters are constantly brought to the surface from the deep ocean.

From greenhouse gases, aerosols, or volcanic activity—can alter the strength and structure of the Walker Circulation by modifying the SST patterns or atmospheric stability. ENSO Events (a result of internal variability but influenced by external forcing): El Niño events weaken the Walker Circulation as warm SST anomalies dominate the central and eastern Pacific; La Niña events strengthen it as cooler SSTs enhance upwelling in the east and convection in the west. Observational studies found the importance of SST in the forcing of atmospheric anomalies.

### 7.1.2 How can surface sea temperature impact atmospheric circulation?

The mechanism by which SST affects the atmosphere primarily involves vorticity balance. Let's consider the vorticity equation, expanding the advective term:

$$\frac{\partial \zeta}{\partial t} = -\nabla \cdot (\zeta + f)\mathbf{V} = -(\zeta + f)\nabla \cdot \mathbf{V} - \mathbf{V} \cdot \nabla(\zeta + f) = -(\zeta + f)D - \mathbf{V} \cdot \nabla(\zeta + f) \quad (7.1)$$

In equation 7.1 D is the divergence and it is equal to

$$D = u_x + v_y = w_z = \frac{\partial w}{\partial z} \quad (7.2)$$

and the last term is the advection.

If we linearize 7.1, all the nonlinear terms ( $D\zeta$ ,  $\mathbf{V} \cdot \nabla\zeta$ ) will be thrown away so that

$$\frac{\partial \zeta}{\partial t} = -fD - v\frac{\partial f}{\partial y} = -fD - \beta v = -f\frac{\partial w}{\partial z} - \beta v \quad (7.3)$$

The equation 7.3 stresses that vorticity evolves due to divergence and meridional velocity. Assuming steady motion

$$f\frac{\partial w}{\partial z} = -\beta v \quad (7.4)$$

NB. Integrating in the vertical this relation, we get

$$-\beta V = f(w_{top} - w_{bottom})$$

for the use off an homogeneous ocean this is the Sverdrup relation, telling us the northward mass transport is equal to the windpress curl divided by beta. In the ocean,  $-\beta v$  is the meridional transport while  $f(w_{top} - w_{bottom})$  is the curl of velocity.

**What about the atmosphere?** We use a shallow water equation centered in the Equator, in beta plane approximation:

$$\begin{aligned}\frac{\partial u}{\partial t} &= -u \frac{\partial u}{\partial x} - v \frac{\partial u}{\partial y} + fv - g \frac{\partial \eta}{\partial x} \\ \frac{\partial v}{\partial t} &= -u \frac{\partial v}{\partial x} - v \frac{\partial v}{\partial y} - fu - g \frac{\partial \eta}{\partial y}\end{aligned}$$

considering a steady motion:

$$u_t - fv = -g \frac{\partial \eta}{\partial x} \quad (7.5)$$

$$v_t + fu = -g \frac{\partial \eta}{\partial y} \quad (7.6)$$

where  $u_t$  and  $v_t$  are the time derivatives. We want to see the system i response to forcing, so we add it and, with  $f = f_0 + \beta y$ ,

$$\begin{aligned}ru - \beta yv &= -g \frac{\partial \eta}{\partial x} + \frac{X}{\rho H} \\ rv + \beta yu &= -g \frac{\partial \eta}{\partial y} + \frac{Y}{\rho H} \\ rg\eta + c^2(u_x + v_y) &= -\frac{gE}{\rho}\end{aligned}$$

where  $X$ ,  $Y$ , and  $E$  are external forcing terms.  $X, Y$  are forcings in  $u, v$ , momentum forcings;  $E$  represents heatings.

The time derivatives are substituted by the dissipation term, representing the effects of nonlinear terms, balancing the equation.

These equations represent a system that is forced steadily that has reached equilibrium, nonlinear terms tend to equilibrate nonlinear instability. Linear problems tend to lead to solutions that grow fast divergently when forced: the linear regime breaks down and nonlinearity takes the command. These equations can be reduced in this form

$$\frac{r}{c^2} (r^2 + f^2) v - r \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right) - \beta \frac{\partial v}{\partial x} = \frac{1}{\rho H} \left[ \frac{r}{c^2} (rY - fX) + r \frac{\partial E}{\partial y} - \frac{\partial}{\partial x} \left( \frac{\partial Y}{\partial x} - \frac{\partial X}{\partial y} + fE \right) \right] \quad (7.7)$$

notice how the left side represents dissipation, while the right holds forcing terms.

If we assume zonally independent motions that are not dependent from  $x$  7.7 becomes

$$\frac{r^2 + f^2}{c^2} v - \frac{\partial^2 v}{\partial y^2} = \frac{1}{\rho H} \left( \frac{\partial E}{\partial y} - \frac{fX}{c^2} + \frac{rY}{c^2} \right) \quad (7.8)$$

The solutions of the free (unforced) equation can be expressed in terms of special function.

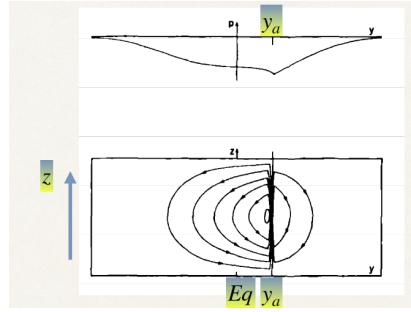


Figure 7.1: Solution of 7.8 eliminating momentum forcing considering heating located at  $y_a$ .

We could solve this equation in the simplest way by eliminating momentum forcing  $(X, Y) = 0$  and considering the heating located at a special latitude  $y_a$  away from the equator, as figure 7.1. This localized heat is controlled by strong events of rains.

As shown in the figure 7.1, the stationary response is asymmetric in longitude and two kinds of waves developed with different shapes: this is due to the fact that the analytical solution, with its specific placed heating source, leads air to rising enhancing compensation mechanism with rain events. The contour lines give the idea of the pressure patterns and there is a motion due to the heat because the upper level of the divergence is compensated by the creation of cycles. The bottom panel shows upper air flow: it's asymmetric in longitude  $E-W$ .

In figure 7.2, as heating  $E$  must be localized, precipitation will have a gaussian-like shape.

However, interesting things happen when the heating varies in  $x$ , and it is localized as the observation of precipitation shows. The precipitation in the mid-latitude is aligned along the preferred path of developing storms, the storm tracks. This precipitation is due to the rising of warm, moist air due to the baroclinic instabilities in the westerly current of the mid-latitudes. Precipitation linked to the baroclinic storm tracks in Winter can be seen from the East coast of North America elongating into Europe and the Mediterranean and from the Asian Far East into the Pacific Ocean. The Summer monsoonal precipitations of the South Asian Monsoon over India and Indochina, of the West Africa Monsoon, and of the weaker North American Monsoon are clearly visible. In general, major convective centers are present over South America and over the Maritime Continent.

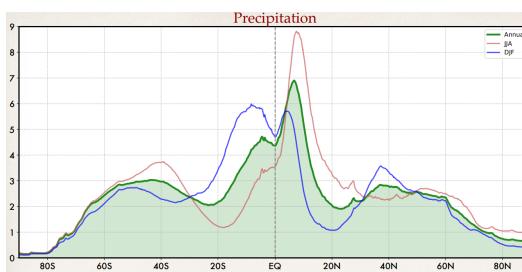


Figure 7.2: Latitudinal distribution of precipitation for Annual mean and Seasonal Cycle. Units are mm/day

An idealized case can be given by

$$E/\rho H = F(x) \exp(-\beta y^2/2c) = F(x) D_0((2\beta/c)^{1/2}y),$$

where  $F$  will be chosen to have the form

$$F(x) = \begin{cases} \cos(\pi x/2L) & \text{for } |x| < L, \\ 0 & \text{for } |x| > L. \end{cases}$$

What about if I assume a basic state, meaning a rest atmosphere, and do the same

operations as before? Imposing specific boundary conditions led us to a kinematic condition  $\frac{Dh}{Dt} = w_{bottom}$  telling us that if air is heating near a mountain, this air does not go inside the mountain but the flow will go above it. This phenomenon stresses that mountains are a steady source of vorticity.

In 1979 Grose and Hoskins [Grose and Hoskins, 1978](#) published a paper describing an investigation that has been conducted of the steady response to orography as described by the linearized shallow-water equations on the sphere. Considering the vorticity equation, but now linearized to respect a basic state  $\bar{u}$  independent of  $x$  and  $F$  is a vorticity source (forcing):

$$\frac{\partial \zeta}{\partial t} = -\bar{u} \frac{\partial \zeta}{\partial x} - f \frac{\partial w}{\partial z} - \beta v + F \quad (7.9)$$

assuming steady motion:

$$0 = -\bar{u} \frac{\partial \zeta}{\partial x} - f \frac{\partial w}{\partial z} - \beta v + F \quad (7.10)$$

a solution can be found, but dissipation in the form Rayleigh damping is needed:

$$-r\zeta = \bar{u} \frac{\partial \zeta}{\partial x} - f \frac{\partial w}{\partial z} - \beta v + F \quad (7.11)$$

Let's now consider this steady forcing with a basic state: barotropic. For an homogeneous case a vertical integration will yield (all quantities vertically integrated):

$$\frac{\partial \zeta}{\partial t} = -\bar{u} \frac{\partial \zeta}{\partial x} - f(w_{top} - w_{bottom}) - \beta v + F \quad (7.12)$$

The rigid lid conditions give  $w_{top} = 0$  and the kinematic condition at the bottom:

$$\frac{Dh}{Dt} = w_{bottom} \quad u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y} = w_{bottom}$$

Linearizing these relation around the basic state we get:

$$-r\zeta = -\bar{u} \frac{\partial \zeta}{\partial x} - f\bar{u} \frac{\partial h}{\partial x} - \beta v + F \quad (7.13)$$

that shows that the mountains are a vorticity source.

Results have been obtained for both idealized and realistic climatological mean zonal flows when perturbed by simple isolated mountains.

$$\begin{aligned} \frac{\partial \zeta}{\partial t} &= -\frac{1}{1-\mu^2} \frac{\partial}{\partial \lambda} (U\zeta) - \frac{\partial}{\partial \mu} (V\zeta), \\ \frac{\partial D}{\partial t} &= \frac{1}{1-\mu^2} \frac{\partial}{\partial \lambda} (V\zeta) - \frac{\partial}{\partial \mu} (U\zeta) - \nabla^2 \left[ \frac{U^2 + V^2}{2(1-\mu^2)} + \phi \right], \\ \frac{\partial \phi}{\partial t} &= -\Phi D - \frac{1}{1-\mu^2} \frac{\partial}{\partial \lambda} [U(\phi - \phi^*)] - \frac{\partial}{\partial \mu} [V(\phi - \phi^*)] \end{aligned}$$

Figure 7.3: Equations on the sphere

where  $\zeta$  is the absolute vorticity,  $D$  the divergence,  $\Phi$  the average height of the fluid multiplied by the gravitational constant  $g$ ,  $\phi$  the deviation of the free surface from that height multiplied by  $g$  and  $\phi^*$  the height of the orography multiplied by  $g$ . The latitude and longitude are denoted by  $\theta$  and  $\lambda$  with  $\mu = \sin \theta$ . The variables  $U$  and  $V$  are respectively the zonal and the meridional velocities multiplied by  $\cos \theta$ .

The two scientists want to transform equations in Figure 7.3 in a steady state, and as we demonstrated, to have it they need forcing (external from the system and not described dynamically) and they have to eliminate the time derivative substituting it with the dissipative term. In particular, the forcing here is represented by  $\phi^*$  because scientists arbitrarily impose it.

### 7.1.3 What is the mechanism by which SST is affecting climate?

Sea Surface Temperature (SST) plays a critical role in shaping long-term climate variability, particularly on timescales beyond weeks (synoptic timescales), where its influence on atmospheric circulation becomes pronounced. Unlike land surfaces, which adjust quickly to temperature changes and lack long-term memory, the ocean's thermal inertia allows SST to maintain a prolonged impact on the climate system (the rate of change of  $T$  is the resulting balance of heat emission). On land, soil moisture provides a limited form of memory, influencing land-atmosphere interactions for up to a year, but the ocean's capacity to modulate heat exchange and atmospheric patterns makes its effect far more persistent and complex.

The mechanism by which SST affects the atmosphere primarily involves vorticity balance, a key dynamic in atmospheric motion. In equatorial regions, SST anomalies act as localized heating sources, driving convection that alters circulation patterns. When heating occurs, air rises, leading to asymmetrical atmospheric responses in the east-west direction. This localized heating produces cyclonic activity in the lower atmosphere and upper-level vortices displaced from the heating source due to atmospheric dynamics. Localized heating from SST causes air to rise, which in turn generates cyclonic and anticyclonic vortices at different atmospheric levels. These processes demonstrate how SST anomalies force changes in atmospheric circulation, influencing phenomena such as tropical cyclones, monsoons, and large-scale weather patterns.

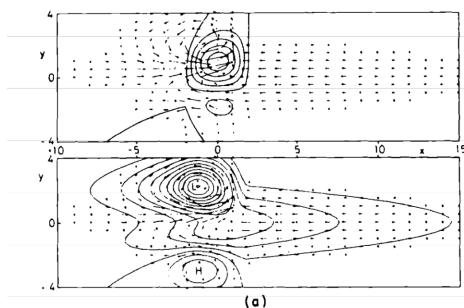


Figure 7.4: Schematic of atmospheric circulation that exhibits cyclonic asymmetry.

In the figure above, the contours represent streamlines or geopotential height isolines in the upper atmosphere, while vectors show wind flow pattern. The cyclonic circulation (low-pressure system "L") indicates counterclockwise motion in the NH. The pattern is asymmetrical meaning that the cyclonic core does not produce perfectly concentric circulation due to asymmetries in forcing. In the high-pressure region "H", the arrows and contours suggest a wave-like structure extending downstream of the low-pressure system, representing possibly Rossby waves generated by heating or perturbation. Note that the generation of vorticity in the lower and upper layers may have opposite signs.

In simplified models of atmospheric circulation, such as a barotropic system with constant wind and dissipation, external forcing like heating or topography alters the basic state. Dissipation, including radiative cooling, ensures that the energy added to the system by these forcings does not lead to instability, allowing for a stationary solution. Remember Grose and Hoskins illustrate how mountains can generate Rossby waves, which are planetary-scale atmospheric waves that propagate and redistribute energy and momentum. These waves,

influenced by both topographic features and SST anomalies, play a significant role in driving global circulation patterns, such as the jet stream and climate anomalies.

The adjustment to forcing is more challenging over oceans than land because of the ocean's slower response time. SST anomalies serve as a sustained forcing mechanism, particularly in tropical regions, driving long-term changes in atmospheric circulation through convection and vorticity dynamics. By generating Rossby waves and other large-scale atmospheric responses, SST anomalies influence both local weather events and global climate patterns. This intricate coupling between ocean and atmosphere underscores the importance of SST in understanding and predicting long-term climate variability.

## 7.2 Steady heating forcing balances

**Hoskins and Karoly** B.J.Hoskins, 1981 (1981) investigated the atmospheric response to stationary heating using a General Circulation Model (GCM) with a realistic structure. Their approach involved simplifying a complex mathematical system by linearizing the model. The focus was on understanding the stationary solutions that arise from this linearized problem.

To explore the response to heating, they used discretization methods (see Chapter 3). This involves breaking down the atmosphere into grid points or using spectral resolution. For example, with a grid resolution of 2 degrees, there are approximately 18000 points (calculated as  $180/2 \times 360/2$ ), and if the model has 9 vertical levels ( $\rightarrow$  Manabe), this results in 162 000 unknowns. Solving such a system requires constructing a matrix  $\mathbf{A}$ , with dimensions based on the degrees of freedom, and then solving the linear equation  $\mathbf{Ax} = \mathbf{f}$ , where  $\mathbf{x}$  represents the variables of interest, and  $\mathbf{f}$  is the forcing vector.

However, directly constructing and solving such a massive matrix is computationally intensive. An alternative approach is to use the model itself, where the prognostic equations are represented as  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ . Adding a heating term  $Q$ , this becomes  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + Q$ , since  $\mathbf{F}(\mathbf{x})$  is nonlinear, linearization is required to solve for perturbations. By introducing a small perturbation  $\mathbf{x}'$  at one grid point:  $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{x}'$ , where  $\mathbf{x}'$  is structured as a special vector (e.g.,  $[1, 0, 0, \dots]$ ), the model's response provides the derivative of  $\mathbf{F}$  that is just the basic state  $\dot{\mathbf{x}} = \dot{\bar{\mathbf{x}}}$ , forming a i-column of  $\mathbf{A}$ , corresponding to the position of 1 in the special perturbation vector. Repeating this process for all degrees of freedom (grid points or spectral coefficients of divergence, vorticity, ... depending on which discretization model you use) constructs the entire matrix.

### 7.2.1 Balancing advection

To consider the heating we need the vorticity equation and the temperature equations both linearized to respect a basic state  $\bar{u}$  independent of  $x$  and  $Q$  is a heating. Assuming a steady motion a solution can be found, but we need dissipation in the form of Rayleigh damping and radiative cooling (instead of vorticity and potential time derivative respectively).

$$r\zeta = -\bar{u}\frac{\partial\zeta}{\partial x} - f\frac{\partial w}{\partial z} - \beta v \quad (7.14)$$

$$\gamma\theta = -\bar{u}\frac{\partial\theta}{\partial x} - v\frac{\partial\bar{\theta}}{\partial y} - w\frac{\partial\theta}{\partial z} + Q \quad (7.15)$$

Let's look at the balances in the thermodynamical equation.

$$\bar{u}\frac{\partial\theta}{\partial x} + v\frac{\partial\bar{\theta}}{\partial y} + w\frac{\partial\bar{\theta}}{\partial z} = Q \quad (7.16)$$

$$f\bar{u}\frac{\partial v}{\partial x} - fv\frac{\partial\bar{u}}{\partial z} + wN^2 = Q \quad (7.17)$$

means:

$$\begin{aligned}
 v &\approx \frac{QH_Q}{f\bar{u}} \quad \text{Zonal Adv} \\
 f\bar{u} \frac{\partial v}{\partial x} - vf \frac{\partial \bar{u}}{\partial z} + w\sqrt{N^2 - Q} &= Q \quad w \approx \frac{Q}{N^2} \quad \text{Vertical Adv} \\
 v &\approx \frac{QH_u}{f\bar{u}} \quad \text{Meridional Adv}
 \end{aligned}$$

If the heating is balanced by the horizontal advection

Figure 7.5: heat balance

In the vertical advection case then the vorticity balance is between the beta terms and the stretching term:

$$\beta v \approx fw_z = \frac{fQ}{N^2 H_Q}$$

The mechanism with smaller  $v$  will dominate, the horizontal advection are of the form:

$$v \approx \frac{QH}{f\bar{u}}$$

and the vertical:

$$v \approx \frac{Q}{N^2}$$

so the dominant process is given by

$$\gamma = \frac{f\bar{u}^2}{\beta N^2 H H_Q} \quad (7.18)$$

Horizontal advection will dominate for small  $H_Q$ , vertical advection otherwise. Depending of the size of  $\gamma$  heating will balanced by horizontal advection if  $\gamma \gg 1$  and by vertical advection if  $\gamma \ll 1$ . In the tropics  $\gamma$  is small if  $H_Q$  is greater than 1 km, so any heating near the surface will generate large meridional motions, but heating away from the surface will generate vertical motions.

$$\begin{aligned}
 \frac{\partial \zeta}{\partial t} &= -\bar{u} \frac{\partial \zeta}{\partial x} - f \frac{\partial w}{\partial z} - \beta v \\
 r\zeta &= -\bar{u} \frac{\partial \zeta}{\partial x} - f \frac{\partial w}{\partial z} - \beta v \quad \text{Stationary solutions}
 \end{aligned}$$

Figure 7.6: Heating and vertical motion

The heating induced by processes like deep convection leads to vertical or horizontal advection, but the dominant mechanism depends on factors such as the stratification of the atmosphere and the strength of the Coriolis parameter. The competition between vertical and horizontal advection can be quantified by a parameter  $\gamma$ , which helps determine which process dominates, the mechanism generating the smaller advection will dominate as it's easier to treat. In equatorial regions, where SST exceeds 28.5°C, strong deep convection generates significant vertical motion, impacting the overall circulation.

### 7.2.2 Balancing vorticity

Heating anomalies, such as those driven by SST, lead to vorticity generation with distinct impacts depending on the region. In the tropics, heating often induces vertical motion and deep convection, while in extratropical regions like the North Atlantic, the vorticity balance is more complex due to baroclinic and barotropic processes. Heating anomalies result in opposite vorticity generation in upper and lower layers, with vertical stratification and horizontal advection playing critical roles.

The implications of anomalous heating differ significantly between regions. For instance, in the North Atlantic, heating anomalies can produce teleconnections, altering large-scale circulation patterns over several months. These teleconnections, resembling the findings of Horace and Wallace on SST variability, highlight how localized heating can influence global atmospheric dynamics.

Teleconnections, or long-distance climate influences, can arise from dynamical causes over intermediate timescales (e.g., months). They are often linked to SST variability and heating-induced circulation changes (e.g., radiative forcings). However, the assumption of a simple atmospheric basic state limits these analyses because real-world atmospheric states are far more complex.

Atmospheric anomalies tend to grow along gradients, where temperature or pressure contrasts enhance the instability and amplify the response to heating. This gradient-based growth underpins many of the patterns observed in teleconnections and regional responses to SST anomalies.

Differences in SST can significantly influence atmospheric processes. Variation in SST can lead to extra convection and extra precipitation (particularly in regions where the ocean surface warms beyond certain thresholds that enhance evaporation and moisture availability). When these responses are systematic and persistent over time, they can act as a form of external forcing on the atmosphere, triggering large-scale atmospheric patterns. These responses can resemble **Rossby waves patterns**<sup>1</sup>.

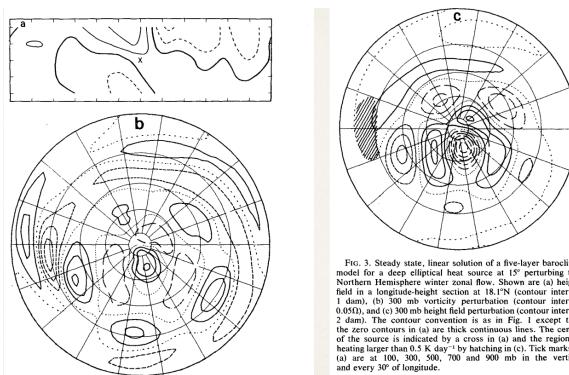


FIG. 3. Steady state, linear solution of a five-layer baroclinic model for a deep elliptical heat source at  $15^{\circ}$  E perturbing the Northern Hemisphere winter zonal flow. Shown are (a) height field in a horizontal section at  $18.75^{\circ}$  N (contour interval 1 dam), (b)  $300\text{ mb}$  vorticity perturbation (contour interval 0.480), and (c)  $300\text{ mb}$  height field (contour interval 2 dam). The coloration is as in Fig. 1 except that the zero contours in (a) are thick continuous lines. The center of the source is indicated by a cross in (a) and the region of heating is indicated by "x" by hatching in (c). Tick marks in (a) are at 100, 300, 500, 700 and 900 m in the vertical and every  $30^{\circ}$  of longitude.

Figure 7.7: systematic SST-induced convection can produce predictable atmospheric wave patterns.

## 7.3 Atmosphere behaviour: linear or non-linear?

A key question in atmospheric science is whether the atmosphere behaves linearly or non-linearly. For synoptic-scale phenomena (timescales of days/weeks), the atmosphere operates

<sup>1</sup>large scale atm waves resulting from conservation of vorticity in the presence of Earth's rotation and varying lat

in in a highly nonlinear regime, with chaotic dynamics dominating the system's behavior. At longer timescales, when averaged over months or seasons, the system can exhibit more predictable patterns. By focusing on time-mean states, the attempt is to determine whether atmospheric responses to SST variability (for instance), which manifest as anomalies, are linear or nonlinear. This remains a topic of active debate, as atmospheric anomalies often exhibit behaviors influenced by both linear responses to forcing and nonlinear interactions within the system.

Rowell (1995) Rowell, 1995 highlighted the potential for increased atmospheric predictability by exploiting SST data. From Lorenz's work, which identified the chaotic nature of atmospheric dynamics and the inherent predictability limits (two weeks for weather forecasting), Rowell built the concept of **potential predictability**, which measures how strongly SST anomalies influence atmospheric conditions ("potential" indicates that this also depends on predictions of anomalous oceanic forcing). The argument is that if SSTs can be accurately predicted over longer periods, such as months, they can serve as a boundary forcing for the atmosphere, enabling predictions of atmospheric anomalies on similar timescales.

Rowell focused on estimating the potential seasonal predictability of the atmosphere by analyzing results from a General Circulation Model ensemble over multiple decades. His work aimed to determine how well atmospheric anomalies could be predicted when SST variability was used as a forcing mechanism.

The atmosphere behaves as a chaotic system and as a consequence, deterministic predictability is limited. However, it is also well known that at many locations, predictability of seasonal weather statistics is also possible. This arises from what may be termed "external" factors that alter the likelihood of residence in atmospheric attractors, enabling probabilistic forecasts to be made of the seasonal mean state, on condition that the external forcing is itself predictable. The primary source of such external forcing at seasonal timescales arises from sea surface temperature (SST) patterns; these can indeed be predicted, either using coupled dynamical models (e.g., Cane et al. 1986; Barnston et al. 1994; Ji et al. 1994) or statistical models (e.g., Ward et al. 1993; Barnston et al. 1994). It is clearly important to be able to assess where on the globe atmospheric variations are sufficiently affected by oceanic forcing to enable practical seasonal prediction. This requires measurements of atmospheric potential predictability.

## 7.4 Potential predictability

The atmosphere-ocean system exhibits two main types of variability:

1. **Internal variability**, arising from the internal dynamics of the coupled system as a fluid system, such as turbulence, convection, and other processes within the fluid dynamics framework.
2. **External variability**, which is the response of the system to external forcing, such as changes in SST, solar radiation, or anthropogenic influences.

In the context of SST-driven atmospheric anomalies, SST serves as an external forcing mechanism, potentially providing a source of predictability. This raises the question: How strongly does SST variability influence atmospheric anomalies? The strength of this influence determines the degree of potential predictability.

If the control exerted by SST on atmospheric anomalies is strong, then accurate SST predictions would lead to a high degree of atmospheric predictability. This concept can be broken into two components:

1. Internal response: the portion of atmospheric variability that would occur regardless of whether SST varies or remains constant.
2. External response: the portion of atmospheric variability driven directly by SST changes over time.

With the understanding these two components, science can evaluate the relative influence of SST as a boundary forcing mechanism and check the limits of atmospheric predictability in the context of climate modeling.

Rowell's study confirmed that SST anomalies have a significant impact on seasonal atmospheric predictability. It highlighted that large-scale, systematic SST patterns (e.g. those associated with El Niño-Southern Oscillation, or ENSO) create atmospheric responses that can dominate the external variability signal. SST-driven predictability varies across regions and seasons:

- tropical regions show higher predictability due to strong ocean-atm coupling, particularly in areas influenced by ENSO events.
- Mid-latitudes have weaker and less consistent predictability because atm dynamics in these regions are dominated by internal variability

Hence, accurate SST predictions could allow for better forecasting of atmospheric anomalies, especially in regions where SST strongly influences weather patterns.

Potential variability examines how much of the total variance in a climate model can be explained by external forcing versus the system's intrinsic chaotic behavior. If the system is insensitive to external forcing, such as prescribed SST, even perfect SST predictions would not yield meaningful atmospheric predictions. Therefore, a robust analysis is required to separate and quantify these components. A statistical technique, ANOVA is used to partition the variance in the system into components associated with external forcing and internal variability. By analyzing the ensemble simulations, ANOVA helps determine the sensitivity of the atmosphere to SST forcing and distinguishes between variance arising from the chaotic nature of the system and variance driven by external boundary conditions.

#### 7.4.1 ANOVA

The ensemble simulations will provide time series of data in  $(n + 2)$  dimension, if  $n$  is the dimension of the geographical labels. For instance, considering atmospheric variables at a single level will result in data labeled by (latitude, longitude, time, ensemble index). The data can be further aggregated using averages to the level necessary for the analysis, in most cases monthly means are the primary data sets. The objective is to separate the total variance  $\sigma_{TOT}^2$  in the component caused by the external forcing (SST) and the system at rest:

$$\sigma_{TOT}^2 = \sigma_{SST}^2 + \sigma_{INT}^2$$

Imagine that data are available from an ensemble of  $n$  climate simulations, each of  $N$  years in length, each forced by the same varying SSTs and sea ice, and each one differing from the others only by its initial atmospheric conditions. Consider a characteristic,  $X$ , which can be quantified for each individual season, for example, the time mean of some variable at a point, a higher moment statistic, a regional index, or an EOF time coefficient.

$$x_{ij} = \mu_i + \epsilon_{ij}$$

By averaging across ensemble members for each grid point, high-frequency fluctuations are smoothed out. This process retains only the components of variability common across ensembles, isolating the SST-forced response. At each grid point, the average of the ensemble realizations represents the SST-forced mean,  $\mu_i$ , which varies across grid points. The variance of the ensemble mean over time reflects the contribution of SST variability to the total system variability.

To explore how the climate system responds to external forcing and check variability, ensemble simulations are critical, and they can be done in different ways:

1. with the same SST forcing: multiple ensemble are run with identical prescribed SSTs to analyze the system's response. While all simulations "see" the same SSTs, their responses will differ due to the chaotic nature of the atmosphere. These differences help isolate internal variability.
2. with different initial conditions: each ensemble starts with different i.c., leading to divergent evolutions over time even if they are subject to the same SST boundary conditions.
3. with separated internal and external variability

Using a single model for these ensembles relies heavily on the model's accuracy and assumptions. Errors or biases inherent in the model could compromise the reliability of the analysis. To mitigate this, multi-model ensembles are increasingly used. These involve simulations from various climate models, reducing the dependency on any single model's characteristics and providing a more robust estimate of externally forced variability.

If the atmospheric response to SST forcing is weak or highly chaotic, predictability is inherently limited, even with perfect knowledge of future SSTs. Conversely, a strong and consistent atmospheric response to SSTs suggests higher potential predictability, emphasizing the importance of accurately predicting SSTs for reliable seasonal or interannual atmospheric forecasts.

The exploration of ensemble simulations, combined with tools like ANOVA, forms the basis for understanding the sensitivity of the climate system to external forcing and quantifying the relative contributions of internal and external variability. By systematically prescribing SSTs and analyzing the spread of ensemble responses, researchers can assess the potential predictability of atmospheric phenomena while addressing the chaotic nature of the system.

The idea is the following: two ensembles

- A statistical technique ANOVA ("analysis of variance")
- Exploit sensitivity of the climate system to generate ensembles of simulation with the same "external" forcing, i.e. prescribed SST.
- Relies on skill on a single model
- late, use many models

The ratio of SST variance to total variance provides a measure of potential predictability, hence indicates the system's sensitivity to SST variations. If the ratio approaches 1, most of the system's variability is controlled by SST forcing.

$$\rho = \frac{\sigma_{SST}^2}{\sigma_{TOT}^2} \quad (7.19)$$

### Predictability across regions

- Tropical regions: SST has strong control over tropical precipitation variability. The ensemble mean explains a significant portion of the variance, indicating that SST-forced variability dominates. For example tropical precipitation anomalies respond robustly to SST changes, and this relationship is crucial for understanding heat distribution in the climate system.
- Mid-latitudes: the relationship between SST and precipitation variability becomes weaker, and the system exhibits greater internal variability. Heat anomalies driven by precipitation in the tropics can influence mid-latitudes, as seen in the Hoskins and Currie results. However, the predictability is less robust in these regions.

#### 7.4.2 Challenges in observations

Observations provide only one realization of the climate system, unlike models that allow for ensemble simulations. This limitation makes it challenging to disentangle forced variability from internal variability directly. Monthly time series of observed SST can provide a proxy for estimating SST-forced variability but lack the direct ability to separate components of variance as in ensemble modeling.

Potential predictability can be visualized as the fraction of variance explained by SST across all grid points, normalized to range between 0 and 1. This measure shows 7.8: strong potential predictability in the tropics, where SST controls much of the variance in upper-layer variables. Weak potential predictability in mid-latitudes, reflecting greater internal variability and weaker SST control.

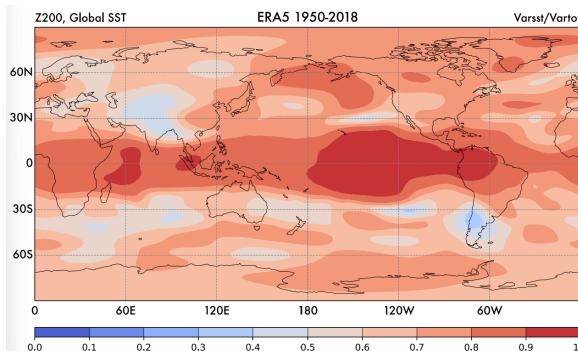


Figure 7.8: Potential predictability as the fraction of variance explained by SST across all grid points.

In mid-latitudes variability is harder to predict due to a weaker direct SST influence and a more dominant role of internal dynamics. Monsoonal circulations while influenced by SST variability, they are not entirely governed by it. Their predictability is thus less robust than tropical SST-forced responses.

Understanding the relative contributions of SST and internal variability is critical for seasonal climate predictions. The strong control of tropical SST over precipitation anomalies offers promising avenues for prediction, especially in the tropics, whereas mid-latitudes remain more uncertain.

Using ANOVA and ensemble simulations, researchers can assess the sensitivity of the climate system to SST forcing. The ratio of SST variance to total variance serves as a measure of potential predictability, which is higher in the tropics due to the strong SST-precipitation coupling and decreases in mid-latitudes due to greater internal variability. This

understanding is foundational for improving seasonal climate predictions and exploring the dynamics of coupled ocean-atmosphere systems.

# Chapter 8

## Earth System Models

Last updated: 2024-12-13

Source file: [chapter-earth-system-models.tex](#)

Earth System Models have evolved from basic atmospheric simulations to complex, integrated frameworks that include oceans, land, vegetation, and greenhouse gases. These models are essential for understanding the nonlinear dynamics of climate change and for making reliable projections about future conditions. By incorporating both natural and anthropogenic forcings, ESMs provide a robust tool for addressing pressing questions about climate variability and potential future scenarios.

Until recently, climate studies predominantly relied on atmospheric models, with only preliminary experiments involving ocean models. However, the significant influence of sea surface temperature (SST) on atmospheric processes necessitated the integration of ocean models with atmospheric models. This coupling led to the emergence of Earth System Models (ESMs), which aim to simulate the dynamic interactions between various components of the Earth system. Early models considered vegetation and certain atmospheric gases as external forcings. However, as climate science advanced, external forcings were redefined to include anthropogenic influences like greenhouse gas concentrations, while other previously external factors, such as ocean dynamics, became integral components of the coupled model. For example,  $CO_2$  is well-mixed in the atmosphere, meaning its concentration is nearly uniform globally.  $CO_2$  is a major contributor to radiative forcing, with approximately 58% of the radiative effect attributed to it. Quantifying the rise in  $CO_2$  requires models capable of simulating its interactions with the ocean, land, and atmosphere. In this model SST should not be considered as external forcing, that now becomes the  $CO_2$ .

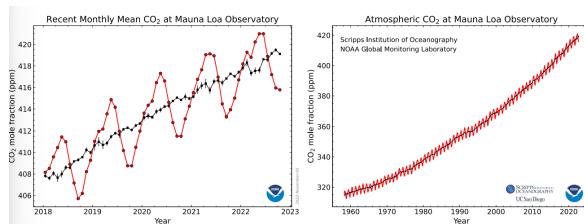


Figure 8.1: Monthly mean  $CO_2$  at Mauna Loa

Methane ( $CH_4$ ) emissions are increasing and are measured in parts per billion. Unlike  $CO_2$ , which is primarily emitted from combustion processes,  $CH_4$  emissions arise from livestock (e.g., cattle farming), agricultural practices (e.g., rice paddies), industrial activities, including leaks in extraction industries. Ice cores provide critical data on past climate oscillations, revealing historical  $CO_2$  and  $CH_4$  concentrations. These records show the natural variability of greenhouse gases over millennia, serving as a baseline for assessing the impacts of recent anthropogenic emissions.

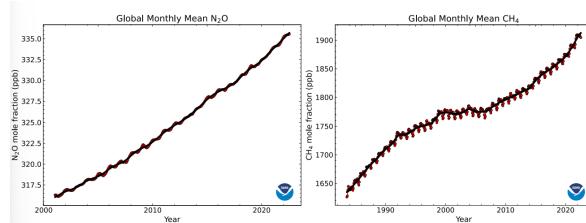


Figure 8.2: Emissions of  $CH_4$  in part per billion.

To project future climate conditions, models incorporate assumptions about socioeconomic scenarios that dictate greenhouse gas emissions. Stone and WeaverStone and Weaver, 2000 work highlighted that climate change would not merely shift temperature distributions but would fundamentally alter their characteristics, leading to nonlinear alterations, underscoring the need for comprehensive modeling frameworks.

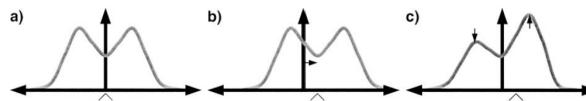


FIG. 1. Comparison of linear and nonlinear interpretations of the projection of climate change. Suppose a hypothetical mode of climate variability has the multimodal PDF shown in (a). Climate change could project (b) linearly onto this mode through a translation of the PDF, or (c) nonlinearly through a change in the shape of the PDF, in this case a shift in the residence frequency of the two regimes associated with this mode. Note that both types of projection result in shifts in the mean of the PDF, indicated by the chevron.

Figure 8.3: Comparison on linear and nonlinear interpretations of the projection of climate change.

In the fig8.3 a) represents a situation where warm and cold temperatures are equiprobables, b) redefines what is cold or warm, c) climate change will displace much more warmer than colder over the year.

## 8.1 Expectations with increasing greenhouse gases

Climate projections often begin with control runs:

1. initialization: start with current atmospheric  $CO_2$  levels. This is also referred to as *control run* (reference experiment). A baseline simulation under current climate conditions, ensuring initial energy balance in the system.
2. Incremental forcing: increase  $CO_2$  concentration by 1% annually until it doubles or quadruples relative to the baseline ( $2xCO_2$ ,  $4xCO_4$ ). This is referred to as a *perturbation experiment*, introducing changes and allowing the system to evolve until the new equilibrium is reached, tracking the stabilization process, ensuring that energy balances across the system.
3. Equilibrium simulation: run the model until a stable equilibrium is reached, ensuring the ocean and atmosphere are in balance: the system must stabilize energetically. This avoids "drift," where the model might otherwise converge toward unrealistic or biased states. The atmosphere typically stabilizes more quickly, while ocean circulation requires longer timescales, often 4 years to stabilize circulation.

Predictive models require accurate initial conditions obtained from observations via constrained estimation methods "data assimilation". Climate models require equilibrated atmosphere and ocean, therefore a spin-up period is usually used. Spin-up is the time taken for an

ocean/atmosphere model to reach a state of statistical equilibrium under the applied forcing.

Increasing greenhouse gas concentrations enhance atmospheric opacity, trapping more longwave radiation. This effect leads to a rise in surface temperatures, what about the other variables?

Changes beyond surface warming will change the character of climatic fluctuations, not just the mean temperature (panel c) of 8.3. This includes shifts in the distribution and variability of other variables, such as precipitation and atmospheric circulation.

Analyzing changes in variability is crucial. For example, EOF analysis is used to identify dominant modes of variability within the control integration. This helps differentiate natural variability from anthropogenic impacts.

**Characteristics of warming** Warming is more pronounced in polar regions due to feedback mechanisms like ice-albedo interactions and changes in heat transport. Land areas warm faster than oceans due to lower heat capacity and less mixing compared to water bodies. Oceans act as heat sinks, absorbing significant portions of excess energy. However, this occurs over longer timescales, delaying the full extent of warming. Surface warming occurs rapidly on land and polar regions, while oceanic warming unfolds over decades or longer.

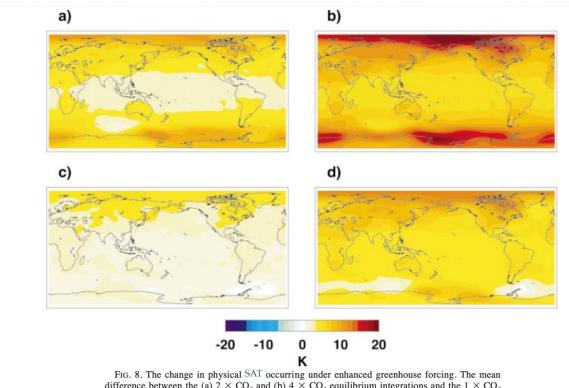


FIG. 8. The change in physical SAT occurring under enhanced greenhouse forcing. The mean difference between the (a)  $2 \times \text{CO}_2$  and (b)  $4 \times \text{CO}_2$  equilibrium integrations and the  $1 \times \text{CO}_2$  control integration is shown, along with the total trend in the (c)  $1\text{--}2 \times \text{CO}_2$  and (d)  $1\text{--}4 \times \text{CO}_2$  transient integrations.

Figure 8.4: One of the first evidence to the distribution of warming.

Model outputs often present:

- Control run variability: baseline distributions of temperature, precipitation and other variables to compare against perturbed states.
- Amplification and shifts: evidence of stronger warming trends over polar regions and continents
- Ocean response: visual representations of delayed oceanic warming compared to surface or land responses.

## 8.2 Expanding the complexity of climate modeling

**Pioneering exercises in Earth System Modeling** If climate characteristics are changing, it means that the characteristics of vegetation will change. For example, 15 vegetation classifications emerge based on temperature and precipitation. As climate changes, vegetation responses must be integrated into the model.

Particles, such as aerosols, also influence temperature and climate. Their effects depend on particle size, with smaller aerosols increasing albedo. Modeling aerosol effects requires

intricate calculations and includes microphysics, where microscopic processes play a significant role.

Coupling components at higher frequencies—shifting from daily to hourly exchanges between Earth-atmosphere and ocean-atmosphere—became necessary for accurate simulations.

**Equilibrium climate sensitivity** Equilibrium Climate Sensitivity (ECS) is historically estimated at 2.2–2.5°C, ECS measures the temperature response to CO<sub>2</sub> doubling. ECS may evolve with baseline temperatures, complicating future predictions (if climate sensitivity depends on basic temperature, increasing it doesn't mean that climate sensitivity will stay the same). However, as baseline temperatures rise, ECS may also evolve, making linear assumptions inadequate. Modern climate projections, driven by socioeconomic scenarios, aim to quantify these nonlinear changes, such as shifts in precipitation patterns (15–20% increase) and polar amplification of warming.

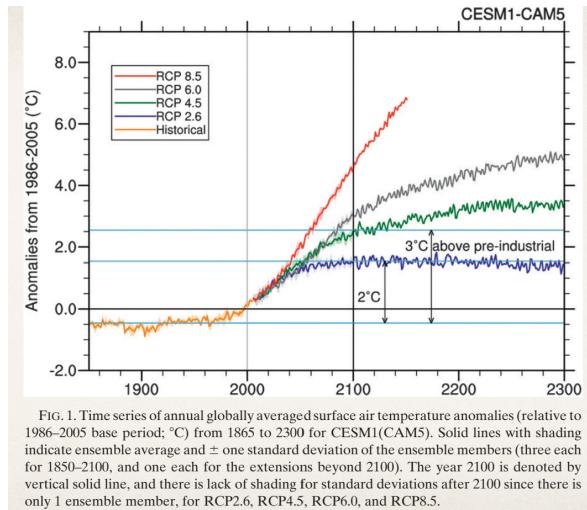


FIG. 1. Time series of annual globally averaged surface air temperature anomalies (relative to 1986–2005 base period; °C) from 1865 to 2300 for CESM1(CAM5). Solid lines with shading indicate ensemble average and  $\pm$  one standard deviation of the ensemble members (three each for 1850–2100, and one each for the extensions beyond 2100). The year 2100 is denoted by vertical solid line, and there is lack of shading for standard deviations after 2100 since there is only 1 ensemble member, for RCP2.6, RCP4.5, RCP6.0, and RCP8.5.

Figure 8.5: Climate scenarios including social economics assumptions

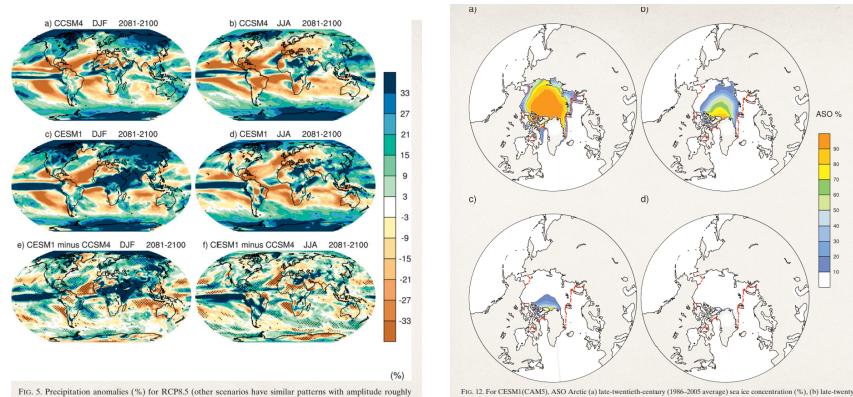


Figure 8.6: Precipitation anomalies (%) and sea ice concentrations (%)

### 8.3 Carbon and Biogeochemical cycles

Some definitions:

- *Geochemistry:* studies the chemical composition of the Earth and the cycling of its chemical elements and major compounds that circulate between the natural reservoirs of the outer envelope lithosphere, atmosphere, hydrosphere and biosphere.
- *Biogeochemistry:* studies the effects of living organisms on chemical composition and cycling.
- *Biogeochemical cycles:* studies the transport and cyclic transformation of chemical elements (or compounds) influenced by the activity of living organisms.

**Chemical composition of the ocean** The chemical composition of the ocean is driven by air-sea exchange: dissolution in air or in surface waters (where solubility is given by the products of partial pressure and the solubility coefficient, which is different for each gas and de-(in-)creases with in-(de-)creasing water temperature); by atmospheric deposition; by accumulation of terrestrial inputs and by removal through marine sediments.

### 8.3.1 The Carbon cycle

Carbon is the defining element of organic matter: virtually all organisms are based predominantly on carbon compounds, it is present in all reservoirs: biosphere, hydrosphere, atmosphere and lithosphere. Atmospheric components  $\text{CO}_2$  and methane ( $\text{CH}_4$ ) are greenhouse-gases, the latter being less abundant ( $\tilde{100x}$ ), but more effective ( $20x$ ), but also unstable (turnover time  $\tilde{10y}$ ). Carbon is also found in transformed sedimentary rocks from marine

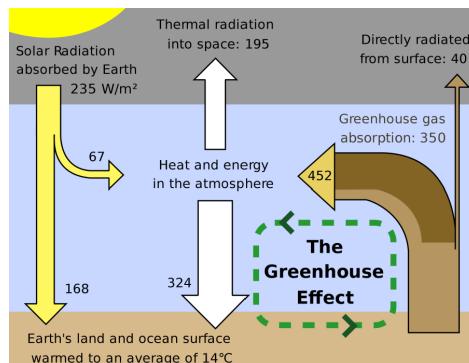


Figure 8.7: Greenhouse effect affects radiative balance. Greenhouse gases act as external forcing but what we are really interested in are emissions: what percentage of  $\text{CO}_2$  stays as concentration and doesn't go away in the biological system? What part of these will be absorbed, which goes in the biochemical system and so on. A certain portion of the emission will be digested by the system, the extra will go into the atmospheric concentration, the net result is to have a model much more difficult to understand and model.

sediments. Inorganic forms in the ocean are separated into dissolved (DIC) and particulate (PIC), including calcareous structural material originating from plankton, corals and shells.

The carbon cycle, an integral part of Earth System Models (ESMs), includes fluxes among the atmosphere, ocean, and land. Carbon dioxide absorption occurs until equilibrium between the atmosphere and ocean is reached. Processes like sediment removal also play roles in long-term carbon storage. Vegetation acts as both a source and a sink for  $\text{CO}_2$ , dynamically interacting with atmospheric concentrations.

The interplay of carbon with other cycles (e.g., nitrogen and phosphorus) further complicates these models. The equations governing these systems combine advection, diffusion, and

biogeochemical reactions, often expressed as:

$$\frac{\partial C}{\partial t} = \mathbf{v} \cdot \nabla C = Sources - Sinks$$

**Oxygen cycle** Oxygen is the most abundant element in the Earth's crust and second most abundant in the ocean. Its presence is due to imbalance of photosynthesis and respiration, guaranteed by the deposition of organic compounds in sediments, where they can not be oxidised. Oxygen is toxic to anaerobic organisms and at high concentrations even to aerobic ones. Cyanobacteria were the first organism to perform photosynthesis, in the "Great Oxygenation Event". Oxygen is produced by the splitting of the water molecules in photosynthesis, approximately half of this production occurs in the ocean. Dissolved oxygen supply is high in the surface ocean due to air-sea exchange and photosynthesis. It is reduced in the intermediate layer due to consumption, particularly at low to mid latitudes. It generally increases again in the deep ocean where it is supplied by deep water formation and barely consumed.

Subtropical gyres, vast regions of rotating ocean currents, are often described as the deserts of the ocean due to their low biological productivity. These gyres are characterized by warm, nutrient-poor waters that limit phytoplankton growth and, in turn, reduce the primary production essential for sustaining marine food webs. Despite their low productivity, these regions play a significant role in global biogeochemical cycles by acting as carbon sinks, absorbing atmospheric CO<sub>2</sub> and sequestering it in the ocean interior over long timescales.

Ocean acidification, a direct consequence of increasing CO<sub>2</sub> concentrations, poses a profound threat to marine ecosystems, including those in subtropical gyres. As the ocean absorbs more CO<sub>2</sub>, it forms carbonic acid, lowering the pH of seawater and increasing its acidity. This shift affects the carbonate chemistry of the ocean, impairing the ability of calcifying organisms, such as corals and certain plankton species, to build and maintain their calcium carbonate structures. Warm-water corals, for instance, are particularly vulnerable; when combined with rising temperatures, acidification leads to widespread coral bleaching and mortality.

These changes are not uniform and depend on regional factors, but the overall trend toward more acidic and nutrient-depleted subtropical gyres highlights the intertwined impacts of climate change on ocean dynamics and marine life. Understanding these processes is essential for predicting the future health of ocean ecosystems and their role in regulating Earth's climate.

There's a wide range of approaches to the marine ecosystem models, such as eulerian/lagrangian modelling, individual based models, food-web models, dynamic size spectrum models and so on.

We can write a more detailed dynamic equation for biogeochemical tracers, as the change of biogeochemical concentrations in the ocean interior is given by: advection, diffusion (turbulent and molecular, the latter is usually neglectable), biogeochemical sources and sinks (production, losses, vertical movements):

$$\frac{\partial C}{\partial t} = -\mathbf{v} \cdot \nabla C + K\nabla^2 C + \frac{\partial C}{\partial t}|_{bgc} \quad (8.1)$$

in its simplest form, biogeochemical changes are assumed proportional to the tracer concentration, where  $r$  may be a function of other state variables:

$$\frac{\partial C}{\partial t}|_{bgc} = rC$$

An **Earth System Model (ESM)** needs to take into account atmospheric dynamics, ocean dynamics, cryosphere, hydrography, atmospheric chemistry, terrestrial ecosystem, marine ecosystem.

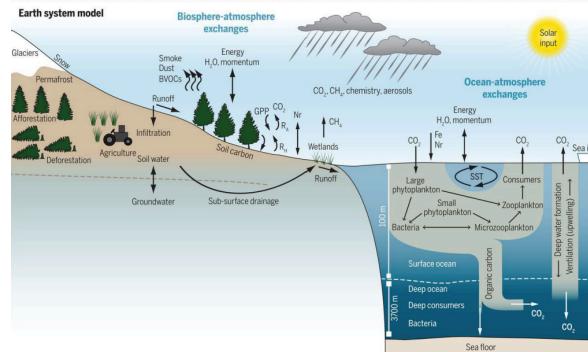


Figure 8.8: Earth system model

### 8.3.2 Complexity and model evaluation

Advancements in ESMs have dramatically increased complexity due to:

1. Higher resolution (spatial and temporal, horizontally and vertically), enhancing model accuracy but demanding greater computational resources: more variables to compute, but increasing resolutions means more degree of freedom.
2. Inclusion of biogeochemical tracers for carbon, nitrogen, and phosphorus cycles.
3. Expanding interactions among components like sea ice, topography, and biological systems.

Model performance is evaluated against satellite observations, revealing improved simulation fidelity over time, especially in capturing chlorophyll distributions and carbon fluxes.

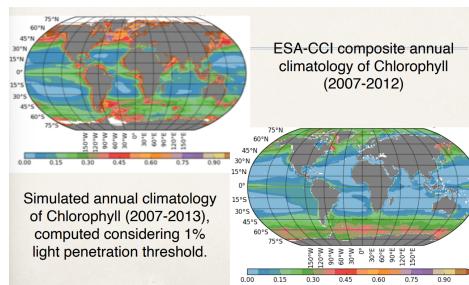


Figure 8.9: How good are these models, from satellites (left) and model (right).

Efforts like the CMIP6 project coordinate global climate scenarios, extending from pre-industrial times (850 CE) to future projections (2100 CE). These simulations consider volcanic eruptions, orbital variations, historical land-use changes. Models predict that both land and ocean systems, traditionally acting as carbon sinks, may transition into sources as emissions continue. These insights highlight the necessity of understanding and mitigating anthropogenic impacts on Earth's climate systems.

In climate system modeling, one of the primary tools is the 3D model that provides a comprehensive representation of various environmental systems. This model allows us to observe what happens in different regions of the Earth, including complex areas like the deep ocean. A challenge in working with these models is identifying patterns, as each component—whether

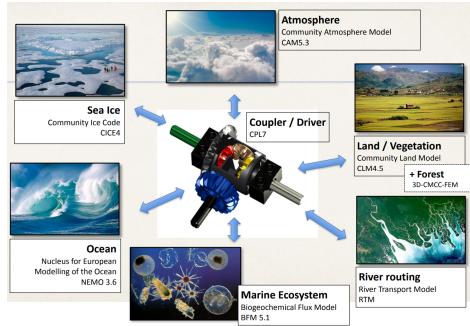
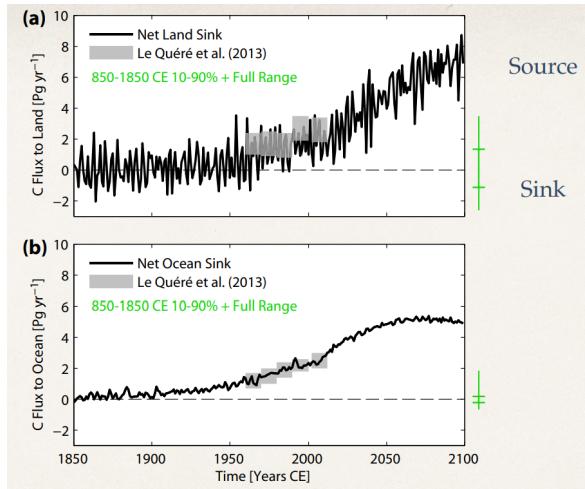


Figure 8.10: CMCC Earth System model for CMIP6



**Figure 4.** Annual mean net carbon flux from the atmosphere to (a) land and (b) ocean. Green bars given the full and 10–90 % range from the preindustrial part of the simulation. Observational estimates are from Le Quéré et al. (2013).

Figure 8.11: Fluxes of carbon between ocean and land, if above 0 → source. In this model both land and ocean become sources of carbon fluxes

species, ecosystems, or regions—follows unique behavior or patterns. These patterns can vary significantly, making it difficult to predict or generalize. However, once these patterns are understood, we can better strategize interventions or "fixes" to mitigate negative effects.

Since climate models often involve simulating multiple scenarios based on different assumptions and variables, it's essential to consider a variety of possible outcomes. One key measure in these simulations is radiative forcing, a value that quantifies the energy imbalance in the climate system caused by factors like greenhouse gas emissions. The radiative forcing value for each scenario is represented by the last two digits (e.g., **0.36**, **0.41**, etc.), which correspond to specific scenarios or pathways of emissions. Even though many combinations of factors may lead to the same radiative forcing value, understanding how these different scenarios play out is crucial in assessing their impacts.

A significant factor in determining the outcomes of climate models is the amount of **greenhouse gases (GHGs)** in the atmosphere. The more GHGs present, the stronger the overall climate response. For example, the **RCP 2.6** scenario, which involves significant emission reductions, requires **negative emissions** (i.e., removing carbon from the atmosphere) after 2050–2060 in order to limit global temperature increases to **2°C**.

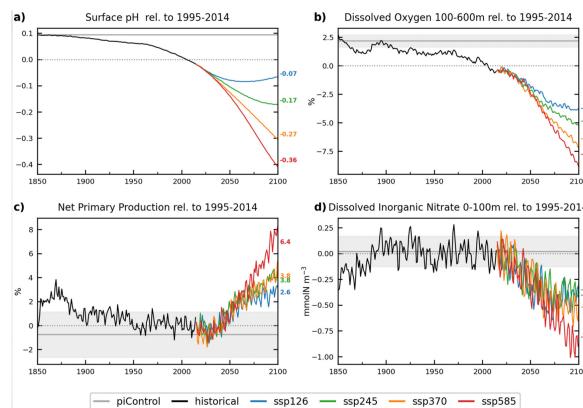


Figure 8.12: Models and predictions

One of the direct consequences of global warming is the increase in temperature, which affects various oceanic processes. For example, an increase in temperature can lead to a decrease in dissolved oxygen levels in the ocean. Warmer waters can hold less oxygen, which has serious implications for marine life. Additionally, nitrate levels are expected to decline due to changes in the ocean's biological production. The reduction in nutrient availability results in lower oxygen production, further exacerbating the oxygen depletion in marine ecosystems.

**Equilibration and Feedbacks** In equilibration experiments, we simulate the effects of stopping emissions and assess how the system reacts. For instance, if emissions were to stop, the concentration of greenhouse gases in the atmosphere would stabilize, and the system would eventually reach a new equilibrium. However, it takes time for the temperature to stabilize, and there is typically a time lag of 10–15 years before equilibrium is reached. The full impact of warming may take 15–20 years to manifest, reflecting the cumulative effect of past emissions. This delay is akin to a Green's function, where the system's response to a perturbation is integrated over time.

This phenomenon of committed warming means that even if emissions were reduced immediately, the Earth would continue to warm for some time due to the inertia in the climate system. Ensemble runs are often used in these experiments to explore a range of possible outcomes, with each run representing a different set of initial conditions or assumptions.

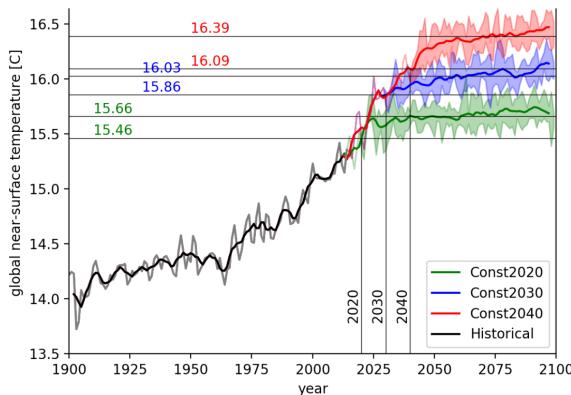


Figure 8.13: Equilibration experiments

## 8.4 Digital Twins

The concept of Digital Twins has gained significant popularity in recent years. A digital twin is a virtual model of a physical system, designed to simulate and predict the behavior of that system. This approach is increasingly used in fields like climate modeling to understand complex processes without needing direct physical experimentation. The idea is to create an interconnected feedback loop where the digital model receives real-time information from the physical system, allowing for better prediction and control.

A *digital model* serves as a mathematical representation of a physical system informed by real-world data and experiments. These models aim to produce predictions that closely align with actual observations, improving forecasting capabilities and providing a more accurate representation of the Earth's systems. A *digital shadow* represents the continuous real-time data flow that informs and updates the model.

In contrast, a digital twin is a completely integrated system, it goes beyond just simulating behavior; it integrates real-time feedback and allows for some level of control over the physical system. However, when it comes to the Earth system, we are still far from achieving a fully integrated digital twin due to the sheer complexity and scale of the planet's processes.

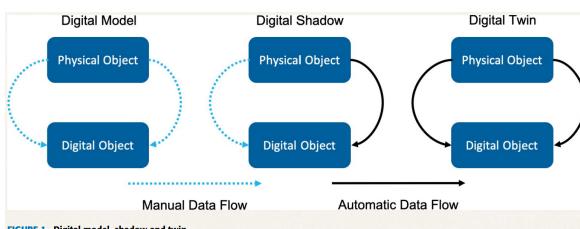


Figure 8.14: Digital world

### 8.4.1 Enabling Technologies

Climate modeling increasingly relies on enabling technologies—tools and innovations that enhance the ability to model complex systems. These technologies might include new ways of gathering observational data, improving the accuracy and resolution of models, or developing better computational methods for simulating climate processes. Unlike basic knowledge, enabling technologies facilitate progress by providing the infrastructure and tools needed to make climate predictions more accurate and actionable.

One focus is the use of non-conventional sources of observation. Traditional data sources like ground-based measurements or satellite observations are valuable but may have limita-

tions. By exploring new types of observations, such as those from new sensors or alternative platforms, we can improve the quality of the models and increase their robustness.

## 8.5 Climate Ensembles

In climate modeling, we work with ensemble models, which involve running simulations under different initial conditions and assumptions. Given that there is inherent uncertainty in predicting the exact future state of the climate, these ensembles help capture the range of possible outcomes.

One of the key challenges is understanding and managing the uncertainty that comes with internal climate variability. While external factors like greenhouse gas emissions are relatively predictable, internal variability (e.g., ocean currents, volcanic activity) is much harder to forecast accurately. This uncertainty affects the predictions, and we typically express it in terms of standard deviation within the ensemble. The spread in the model outputs reflects the uncertainty in predicting specific outcomes, such as temperature rise or the rate of ice sheet melting.

The uncertainty surrounding climate sensitivity—the Earth’s response to changes in greenhouse gas concentrations—is a significant factor in climate projections. Even though we have a general idea of the climate’s response, there are still unknowns, particularly when it comes to regional effects or extreme events. This uncertainty is not just physical, but also depends on the emission scenario chosen. Different emission pathways lead to different climate futures, making the exact trajectory difficult to pin down. Some models predict rapid warming, while others may show slower changes, reflecting this uncertainty.

Ultimately, the challenge of climate modeling goes beyond just predicting specific scenarios. It’s more about understanding **climate risks**—the potential for large-scale, unpredictable changes that could result from different emission scenarios. While we can predict trends like ice sheet melting or temperature increases, the exact timing and scale are much more difficult to determine. The unpredictability of these changes makes climate risk a significant concern, and it highlights the importance of uncertainty management in climate modeling.



# Chapter 9

## Theory of symmetric circulation

Last updated: 2024-12-13

Source file: [chapter-theory-simm-circ.tex](#)

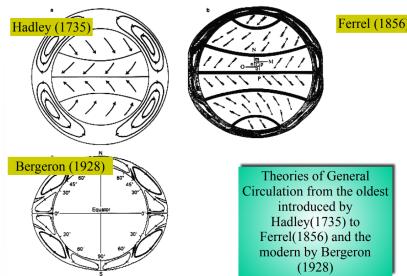


Figure 9.1: History of the theories of the general circulation

In studying the general circulation of the atmosphere, a significant question arises: why does the **Hadley cell**, a key feature of the meridional circulation, extend only to about 30° latitude in each hemisphere? Why doesn't it dominate the entire system, or extend further north or south? To answer these questions, we must delve into the dynamics of the Nonlinear Axially Symmetric CirculationI.Held, 1980 as explored by Isaac Held and others. Held analyzed a simplified atmospheric system by assuming symmetry around the Earth's axis. This eliminates longitudinal variations, allowing a focus on meridional (north-south) and vertical dynamics.

The general circulation of the atmosphere is composed of two major components, as described in the first chapter 1:

1. **Hadley Cell (Direct Circulation):** this thermally-driven circulation dominates the tropics and subtropics. Warm air rises near the Equator, flows poleward at high altitudes, sinks near 30° latitude, and returns equatorward near the surface as the trade winds.
2. **Ferrel Cell (Indirect Circulation):** beyond 30° latitude, the circulation is opposite in direction, driven primarily by midlatitude dynamics rather than direct thermal forcing.

Hadley theorized the existence of trade winds and a direct meridional circulation toward the poles, but his early model incorrectly predicted wind directions in midlatitudes. This error was resolved with the realization that the Earth's rotation introduces Coriolis forces, altering the direction of winds and creating distinct circulation cells like the Ferrel cell.

Then why does the Hadley cell stop ad 30° latitude?

The Hadley cell limit is determined by the interplay of:

- Coriolis Force: as air moves poleward, the Coriolis effect increases with latitude, deflecting the flow and limiting the extent of the Hadley circulation
- Angular momentum conservation: poleward moving air in the upper branch of Hadley cell accelerates as it conserves angular momentum, eventually leading to a breakdown into zonal (east-west) flow.
- Radiative forcing and thermal gradients: the temperature gradient between the Eq and the poles influences the location and strength of the Hadley cell.

To study the symmetric circulation, we simplify the governing equations by assuming:

1. Axial Symmetry: the system is symmetric around the equator, eliminating longitudinal dependencies.
2. Hydrostatic Balance: vertical pressure gradients balance gravity.
3. Conservation of Mass (Continuity) divergence in the meridional plane (latitude-height) is zero.
4. Prescribed Temperature Profile ( $\Theta_E$ ): the system follows a latitudinally dependent potential temperature gradient.

Using these assumptions, the Navier-Stokes equations reduce to a simpler form, looking at the time derivatives:

$$\begin{aligned} 0 &= -\nabla \cdot (\bar{v}u) + fv + \frac{uv \tan \theta}{a} + \frac{\partial}{\partial z} \left( \nu \frac{\partial u}{\partial z} \right) \\ 0 &= -\nabla \cdot (v\bar{v}) - fu - \frac{u^2 \tan \theta}{a} + \frac{\partial}{\partial z} \left( \nu \frac{\partial v}{\partial z} \right) - \frac{1}{a} \frac{\partial \Phi}{\partial \theta} \\ 0 &= -\nabla \cdot (\bar{v}\Theta) - \frac{1}{\tau}(\Theta - \Theta_E) + \frac{\partial}{\partial z} \left( \nu \frac{\partial \Theta}{\partial z} \right) \\ 0 &= -\nabla \cdot (\bar{v}) \\ \frac{\partial \Phi}{\partial z} &= g \frac{\Theta}{\Theta_0} \end{aligned}$$

where  $\bar{v} = (v, w)$  represents the velocity vector in the meridional plane,

$$\nabla = \left[ \frac{1}{a \cos \theta} \frac{\partial \cos \theta}{\partial \theta}, \frac{\partial}{\partial z} \right]$$

is the gradient operator.  $\Theta$  represents potential temperature.  $\Phi$  is the geopotential.  $a$  is the Earth's radius,  $g$  is gravity, and  $\nu$  represents viscosity. These equations are written with the following approximations:

- Steady,  $(y, z)$ , Boussinesq, dry, hemispheric
- Rigid lid at height  $H$
- Forced by a linear "radiative" heating, chosen to be

$$\frac{\Theta_E(\theta, z)}{\Theta_0} = 1 - \frac{2}{3} \Delta_H P_2(\sin \theta) + \Delta_v \frac{z}{H} - \frac{1}{2}$$

where  $\Theta_0$  is a global mean,  $\Delta_H, \Delta_v$  are nondimensional constants that represents the fractional change in temperature from equator to pole and from the top to the bottom, respectively.  $P_2$  is the second Legendre polynomial.

- Vertical diffusion with constant diffusivity
- Zero stress boundary condition at the top and proportional to the surface wind at the bottom:

$$w = 0 \quad \frac{\partial u}{\partial z} = \frac{\partial v}{\partial z} = \frac{\partial \Theta}{\partial z} = 0 \quad \text{at } z = H$$

$$w = 0 \quad \frac{\partial \Theta}{\partial z} = 0, \quad v \frac{\partial u}{\partial z} = Cu, \quad v \frac{\partial v}{\partial z} = Cv \quad \text{at } z = 0$$

The zonal (east-west) wind  $u$  is a function of latitude and height, while the meridional (north-south) wind  $v$  is small or zero. The radiative forcing, structured by latitude and altitude, determines the thermal gradient from equator to pole, driving the circulation.

The equation of motion have a simple solution for the case in which there is no meridional flow ( $v = 0$ ). In this case also the vertical velocity is zero ( $w = 0$ ) and the temperature is in radiative equilibrium ( $\Theta = \Theta_E$ ), the zonal wind satisfies:

$$\frac{\partial}{\partial z} \left( f u_E + \frac{u_E^2 \tan \theta}{a} \right) = - \frac{g}{a \Theta_0} \frac{\partial \Theta_E}{\partial \theta} \quad (9.1)$$

The vertical integration of the momentum equation and using the boundary condition implies that the zonal wind must be zero at the ground, so we can integrate 9.1 and consider the solution obeying  $u = 0$  at  $z = 0$ :

$$\frac{u_E}{\Omega_a} = \left[ \left( 1 + \frac{2Rz}{H} \right)^{1/2} - 1 \right] \cos \theta \quad R = \frac{gH\Delta_H}{\Omega^2 a^2} \quad (9.2)$$

The equations produce nonlinear solutions describing zonal flow. One notable solution shows a zonal westerly wind ( $u > 0$ ) increasing with height and latitude; the dynamics are controlled primary by rotation (Coriolis effect) and temperature gradients, encapsulated in a dimensionless parameter  $R$ , which determines the strength and extent of the circulation.

However, this solution has some difficulty. Considering the zonal momentum equation

$$-\nabla \cdot (\mathbf{v} M) + \frac{\partial}{\partial z} \left( \nu \frac{\partial M}{\partial z} \right) = 0 \quad (9.3)$$

where

$$M = u \times \theta = \Omega a^2 \cos^2 \theta + ua \cos \theta$$

is the angular momentum. We can ask ourselves where it can have a maximum. Hide<sup>1</sup> showed that his maximum cannot be in the interior of the fluid. Assume in fact that such a maximum exist, then integration around a closed circuit sufficiently close to the maximum will get

$$\int \nabla \cdot (\mathbf{v} M) dl \approx M \int \nabla \cdot \mathbf{v} = 0$$

for incompressibility, whereas the integral of the diffusion will not be zero. In practice, a maximum in the interior will be diffused away and the advection will not be capable to counteract. Instead it is possible to have a maximum close to the surface by vertically integrating the momentum equation where it is compensated by the stress boundary condition. The same condition will require that  $u \leq 0$  at the surface. It follows that the angular momentum  $M$  must be always smaller than its equatorial maximum ( $\Omega a^2$ ) everywhere, yielding the condition

$$u < u_M = \frac{\Omega a \sin^2 \theta}{\cos \theta} \quad (9.4)$$

---

<sup>1</sup>Hides theorem

this is the solution that corresponds to the conservation of angular momentum. This means that the radiative solution  $u_E$  cannot be valid until  $u_E > u_M$ . Using  $\Delta_H = 1/3$ ,  $H = 15$  km, we find  $R \approx 0.2$  and therefore

$$\theta_h \leq 27^\circ$$

In this region the meridional velocity  $v$  cannot be zero as required by the radiative balance.

Hence, we saw that if the angular momentum has a maximum inside the fluid, then we can make a contour around it, then integrate the angular momentum equation around it: the integral of the advection part will be zero, the integral of the diffusion will not. So there cannot be a maximum on the interior of the fluid, it will be on the surface, which will be compensated by the boundary condition. The solution  $u_E$  cannot be valid everywhere but in the regions where it is greater than  $u_M$  9.4 that corresponds to a certain latitude interval.

### Problems

The extent of the Hadley circulation, which is currently confined to the tropics and subtropics, is sensitive to changes in the Earth's energy balance and rotation rate. **If climate change alters the equatorial temperature gradient or modifies energy conservation constraints, the Hadley cell could expand or contract, influencing global atmospheric patterns.** The boundary of the Hadley cell, denoted as  $\theta_h$ , depends on key planetary and climatic parameters:

- **Gravity ( $g$ ):** Stronger gravity limits the poleward extent of the Hadley cell, while weaker gravity allows it to expand further north or south.
- **Rotation Rate ( $\Omega$ ):** Faster rotation confines the Hadley cell closer to the equator, while slower rotation enables it to extend poleward.
- **Temperature Gradient ( $\Delta T$ ):** The difference in temperature between the equator and the poles drives the circulation, with larger gradients supporting more extensive cells.

On larger planets with slower rotation rates, the Hadley circulation tends to be smaller and less confined, as planetary rotation plays a weaker role in shaping the circulation.

We can be more precise on the estimate of  $\theta_h$ , i.e., the limit of the Hadley cell. Because a meridional flow must exist within a certain region close to the equator, let's assume that such circulation exists up to a certain latitude  $\theta_h$ . The poleward high altitude branch of the circulation conserves angular momentum so  $u(\theta, H) \approx u_M(\theta)$ , but surface drag at the surface is sufficiently strong that  $u(\theta, 0) \approx 0$ . Poleward of the limiting latitude  $\theta_h$  the circulation is zero and in radiative equilibrium,  $\Theta = \Theta_E$ ,  $u = u_E$ . We can evaluate the difference between the top ( $z = H$ ) and the bottom ( $z = 0$ ) in 9.1:

$$f[u(H) - u(0)] + \frac{\tan \theta}{a} [u^2(H) - u^2(0)] = -\frac{gH}{a\Theta_0} \frac{\partial \{\Theta\}}{\partial \theta} \quad (9.5)$$

where  $\{\Theta\}$  denotes the vertical mean. Inserting

$$f = 2\Omega \sin \theta \quad u(H) = u_M = \frac{\Omega a}{\cos \theta} \sin^2 \theta$$

and integrating from 0 to  $\theta$ , we want to consider the difference between the vertical means:

$$\frac{\{\Theta(0)\} - \{\Theta(H)\}}{\Theta_0} \approx \frac{\Omega^2 a^2 \sin^4 \theta}{2gH \cos^2 \theta} \approx \frac{1}{2} \frac{\Omega^2 a^2}{gH} \theta^4 \quad (9.6)$$

assuming small angles.

We also want to impose the continuity of potential temperature at latitude  $\theta_h$ :

$$\{\Theta(\theta_h)\} = \{\Theta_E(\theta_h)\} \quad (9.7)$$

Combining 9.6 with the 9.7 integrated, that is:

$$\frac{\{\Theta_E\}}{\Theta_0} = \frac{\{\Theta_E(0)\}}{\Theta_0} - \Delta_H \theta^2$$

the solutions are

$$\frac{\{\Theta(0)\}}{\Theta_0} = \frac{\{\Theta_E(0)\}}{\Theta_0} - \frac{5}{18} \Delta_H R \quad \theta_h = \sqrt{\frac{5}{3} R} = \sqrt{\frac{5gH\Delta_H}{3\Omega^2} a^2} \quad (9.8)$$

inserting numerical values we get

$$\theta_h \approx 26^\circ \quad (9.9)$$

This theory that is based on the conservation of angular momentum and of energy (via the conservation of potential temperature) then conclude that the extent of the Hadley cell is proportional to the square root of the equator-pole temperature gradient and to the square root of the height of the troposphere. It is instead inversely proportional to the rotation rate and the radius of the Earth, larger planets or faster planets will get smaller Hadley cells, whereas slower planets will result in larger Hadley cells.

We see how  $\theta_h$  is a function of gravity, rotation of the planet, temperature change from equator to pole ( $\rightarrow$  bigger planets have smaller Hadley circulation cell). Changing rotational rate, how does that relation change? The figure (Figure 9.2) illustrates the impact of decreasing

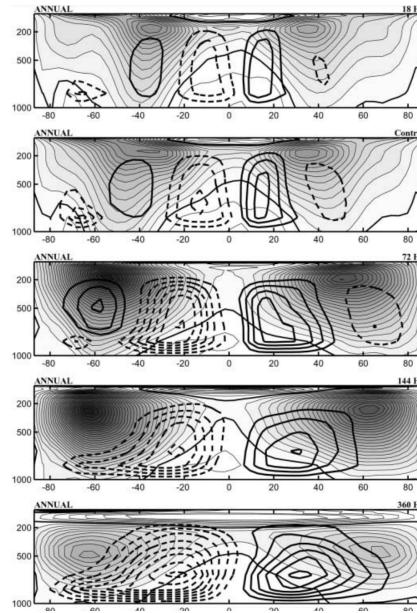


Figure 9.2: decreasing Earth's rotation rate on the Hadley circulation

ing Earth's rotation rate on the Hadley circulation. Thick lines represent the streamfunction, indicating the direction of the flow:

- With Earth rotating at its current rate (24-hour day), the Hadley cell extends to about 30° latitude, with the Ferrel cell occupying the midlatitudes and storm tracks forming over these regions due to transient turbulence.

2. If Earth's rotation slowed to a 72-hour day, the Hadley circulation would expand significantly, shifting its poleward boundary and the jet streams northward.
3. With a rotation period of 360 hours (15 days), the Hadley cell would extend all the way to the poles, effectively eliminating the Ferrel cell.

These simulations use a General Circulation Model (GCM), averaging results over time and incorporating conservation of momentum and energy.

The stability of the Hadley cell depends on how momentum is transported from the equator to the poles. In westerly momentum transport, the Hadley cell drives momentum poleward in the upper troposphere. Ferrel Cell and turbulence: the midlatitudes are dominated by transient eddies, which result from turbulent flow and instabilities.

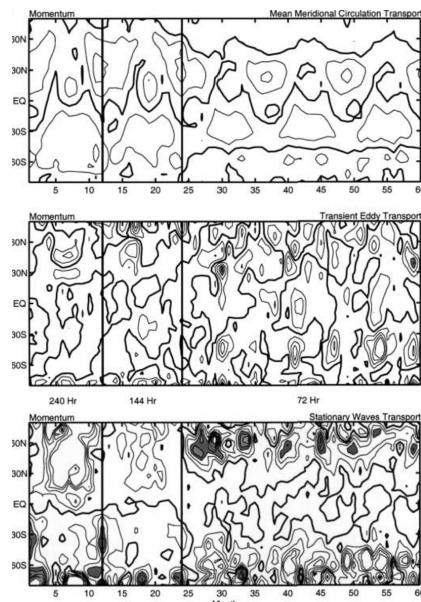


Figure 9.3: Stability of Hadley cell

As Earth's rotation slows (by approximately 1 hour every 100 million years due to tidal interactions), the momentum transport dynamics could change, potentially destabilizing current circulation patterns.

**Precipitation and Storm Tracks** Changes in the Hadley circulation's extent also affect precipitation patterns:

- **Tropics:** Expanded Hadley circulation would shift tropical rainfall belts poleward.
- **Midlatitudes:** Storm tracks, driven by transient turbulence in the Ferrel cell, would move poleward, altering precipitation patterns in these regions.

# Chapter 10

## Tipping points

An ‘abrupt change’ is defined in this report as a change that takes place substantially faster than the rate of change in the recent history of the affected component of a system. In some cases, abrupt change occurs because the system state actually becomes unstable, such that the subsequent rate of change is independent of the forcing. We refer to this class of abrupt change as a ‘tipping point’, defined as a critical threshold beyond which a system reorganizes, often abruptly and/or irreversibly. Tipping points are also referred to as planetary boundaries, they represent thresholds where rapid, often irreversible, changes occur in the climate system. These transitions typically involve fast dynamics combined with irreversibility, and they are often associated with bistable systems, which can exist in two distinct states (potential wells). Transitions between these states can be triggered either by noise-induced perturbations or by modifications of the underlying potential, often referred to as bifurcations.

In the figure 10.1, the ‘valleys’ represent different climate states the system can occupy, with ‘hilltops’ separating the stable states. The resilience of a climate state is implied by the depth of the valley. The current state of the system is represented by a ball. Both scenarios assume that the ball starts in the left-hand valley (dashed-black lines) and then through different mechanisms dependent on the type of tipping transitions to the right-hand valley (colored lines). Noise-induced tipping events (a, b), for instance drought events causing sudden dieback of the Amazon rainforest, develop from fluctuations within the system. The stability landscape in this scenario remains fixed and stationary. A series of perturbations in the same direction, or one large perturbation, are required to force the system over the hilltop and into the alternative stable state. Bifurcation tipping events (c, d), such as a collapse of the thermohaline circulation in the Atlantic Ocean under climate change, occur when a critical level in the forcing is reached. Here the stability landscape is subjected to a change in shape. Under gradual anthropogenic forcing the left-hand valley begins to shallow and eventually vanishes at the tipping point, forcing the system to transition to the right-hand valley.

The present rates of response of many aspects of the climate system are proportionate to the rate of recent temperature change, but some aspects may respond disproportionately. Some climate system components are slow to respond, such as the deep ocean overturning circulation and the ice sheets. It is virtually certain that irreversible, committed change is already underway for the slow-to-respond processes as they come into adjustment for past and present emissions. The paleoclimate record indicates that tipping elements exist in the climate system where processes undergo sudden shifts toward a different sensitivity to forcing, such as during a major deglaciation, where 1°C degree of temperature change might correspond to a large or small ice-sheet mass loss during different stages. For global climate indicators, evidence for abrupt change is limited, but deep ocean warming, acidification and sea level rise are committed to ongoing change for millennia after global surface temperatures initially stabilize and are irreversible on human time scales (very high confidence). At the regional

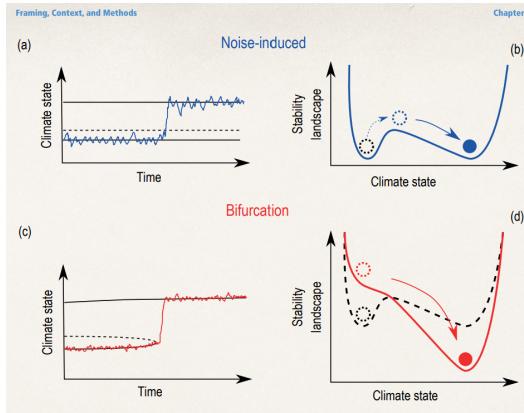


Figure 10.1: Illustration of two types of tipping points: noise-induced (a, b) and bifurcation (c, d). (a) and (c) are example time-series (coloured lines) through the tipping point, with solid-black lines indicating stable climate states (e.g., low or high rainfall) and dashed lines representing the boundary between stable states. (b) and (d) are stability landscapes, which provide an intuitive understanding of the different types of tipping point.

scale, abrupt responses, tipping points and even reversals in the direction of change cannot be excluded (high confidence). Some regional abrupt changes and tipping points could have severe local impacts, such as unprecedented weather, extreme temperatures and increased frequency of droughts and forest fires. Models that exhibit such tipping points are characterized by abrupt changes once the threshold is crossed, and even a return to pre-threshold surface temperatures or to atmospheric carbon dioxide concentrations does not guarantee that the tipping elements return to their pre-threshold state. Monitoring and early warning systems are being put into place to observe tipping elements in the climate system.

## 10.1 Examples of Climate System Tipping Points

1. **Gulf Stream and Heat Transport:** The Gulf Stream, a critical component of the Atlantic Meridional Overturning Circulation (AMOC), is unlikely to vanish entirely. However, its role in heat transport between the tropics and higher latitudes could be altered under changing climatic conditions. Such changes would have profound implications for regional and global climates.
2. **Sea Level Rise:**
  - Thermal Expansion: as oceans warm, their volume increases due to thermal expansion, contributing significantly to sea level rise.
  - Melting Ice Sheets: ice melting currently contributes approximately 3 mm per year to global sea levels.
  - Solid Earth Adjustments: on longer timescales, the solid Earth is rebounding from the removal of massive ice sheets, a process known as isostatic adjustment.
  - Density Variations: sea level is also affected by changes in seawater density. For instance, increased salinity raises density, causing local sea levels to shrink.
3. **ENSO (El Niño–Southern Oscillation):** ENSO's behavior is heavily influenced by the depth of the thermocline in the Pacific Ocean. If processes push the thermocline lower, there is the potential for a **permanent El Niño state**. This has been speculated as a possibility under higher CO<sub>2</sub> concentrations, reminiscent of conditions during the

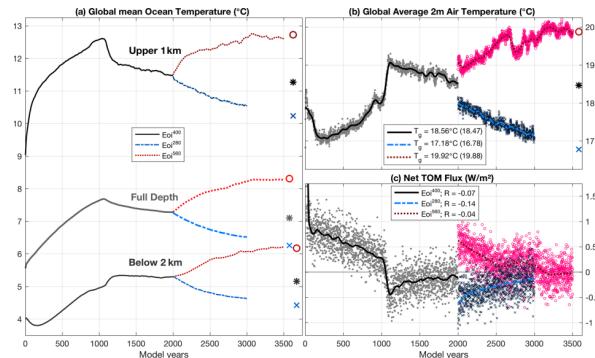


Figure 10.2: Time series of globally averaged temperatures for the entire length of our three different Pliocene simulations: Eoi400 (black), Eoi280 (blue), and Eoi560 (red). Shown are the (a) upper (dark), deep (medium), and full-depth (light) ocean temperature; (b) near-surface air temperature; and (c) globally averaged top-of-model (TOM) net radiative flux. Thick lines in (b) and (c) show the corresponding time series after applying a 100-year smoothing mask. The estimated equilibrium temperatures are indicated at the end in (a) and (b) using large markers and the same color convention. The globally averaged mean temperature ( $T_g$  in b) and net radiative flux ( $R$  in c) over the last 100 years are added in the legends (bracketed values for the estimated equilibrium).

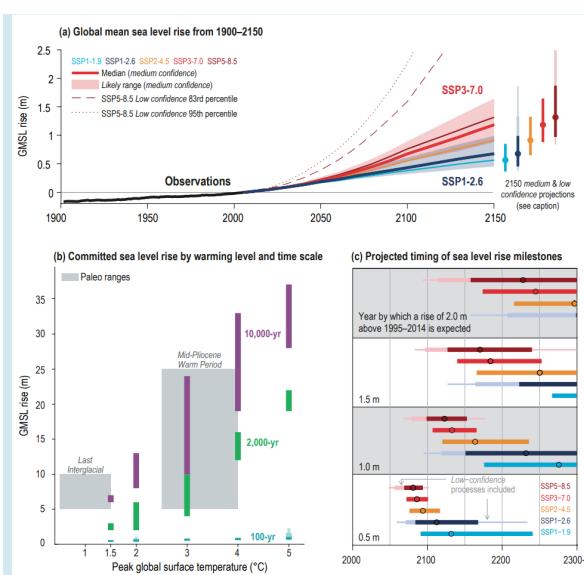


Figure 10.3: Sea level rise

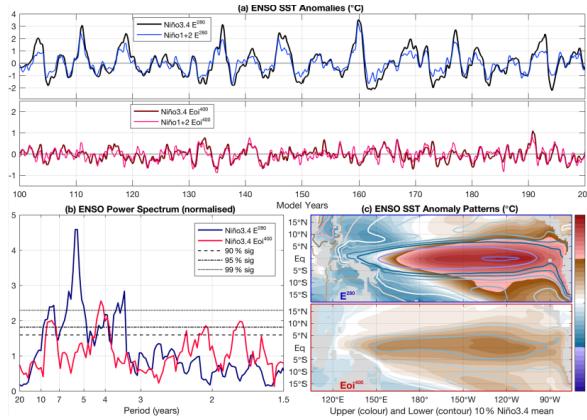


Figure 10.4: (a) ENSO time series for the E280 (blue) and Eoi400 (red) using monthly SST anomaly fields and a 5-month running mean. Note the different scaling. Temperature intervals are kept the same for visual comparison. (b) Multi-taper power spectrum, using 200 years of monthly data, including the 90 %, 95 %, and 99 % confidence levels. (c) Corresponding ENSO SST patterns, taking the mean over the 10% highest (shading) and 10% lowest (contours) monthly Niño 3.4 index values.

Pliocene epoch. However, current experiments are inconclusive regarding this outcome.

4. **Glaciations and Orbital Cycles:** major glaciations have occurred cyclically over the past 3 million years, with a periodicity of approximately 100,000 years. These glaciations are thought to be influenced by orbital forcing, as inferred from isotopic records. However, the exact causal mechanisms remain unclear. While carbon dioxide and glacial cycles covary, the time lag (approximately 10 years for CO<sub>2</sub> effects to manifest) complicates establishing causality.

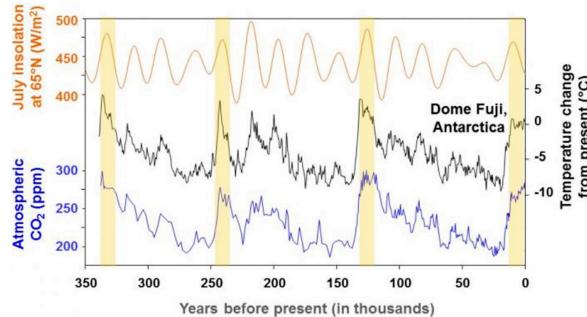


Figure 10.5: Glacial cycles inferred from isotopic records

## 10.2 Climate impacts

1. **Regional Climate Impacts:** A collapse in systems like the AMOC or the establishment of a permanent El Niño state would drastically alter weather patterns, precipitation, and temperature distributions globally.
2. **Sea Level Rise Feedbacks:** Dynamic processes such as shifting currents, salinity changes, and ice melt interact to exacerbate sea level changes, affecting coastal regions worldwide.

3. **Climate Stabilization Challenges:** The path to stabilizing global temperatures involves achieving net-zero carbon emissions. Without this, many tipping points, once crossed, are irreversible.

**Modeling and Understanding Tipping Points** Intermediate-complexity models such as **CLIMBER** are used to explore these phenomena. These models balance detailed physics with computational efficiency, allowing for longer timescales and broader parameter explorations.

**Reversibility and Hysteresis** Studies on hysteresis in climate systems investigate the potential for reversing climate change. While some tipping points (e.g., ice sheet loss) are deemed irreversible, others may exhibit pathways back to previous states under specific conditions. However, reducing east-west gradients in the Pacific through tipping-point dynamics, such as those involving ENSO, remains a critical area of ongoing research. We cannot stabilize the temperature without going to **net 0** emissions.

xx babies



## Chapter 11

# Are General Circulation models obsolete?



# Appendix A

# Appendix

Last updated: 2024-12-13

Source file: [appendixA.tex](#)

## A.1 Spherical harmonics expansion

Spherical harmonics are key mathematical functions in the expansion of scalar and vector fields on the surface of a sphere. In climate modeling, they are extensively used for representing atmospheric and oceanic variables, such as temperature, pressure, and wind patterns, on the spherical surface of Earth.

Let's break down the **spherical harmonics expansion** in more detail, looking at its mathematical formulation, properties, and applications, especially in the context of climate modeling.

### 1. Mathematical Definition of Spherical Harmonics

The spherical harmonics  $Y_l^m(\theta, \phi)$  are solutions to the Laplace equation on the sphere. They depend on two angular coordinates, **latitude** ( $\theta$ ) and **longitude** ( $\phi$ ). The general form of a spherical harmonic is:

$$Y_l^m(\theta, \phi) = N_l^m P_l^m(\cos \theta) e^{im\phi}$$

Where:

- $l$  is the **degree** of the harmonic, a non-negative integer ( $l = 0, 1, 2, 3, \dots$ ).
- $m$  is the **order** of the harmonic, an integer satisfying  $-l \leq m \leq l$
- $P_l^m(\cos \theta)$  are the **associated Legendre polynomials**, which define the angular part of the function in terms of the **colatitude** ( $\theta$ ).
- $e^{im\phi}$  is a complex exponential that encodes the **longitudinal** dependence, where  $m$  controls the azimuthal symmetry.

**Normalization Constant:** The spherical harmonics are typically normalized such that:

$$N_l^m = \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}}$$

This ensures that the spherical harmonics are **orthonormal** over the surface of the sphere. That is:

$$\int_0^\pi \int_0^{2\pi} Y_l^m(\theta, \phi) \overline{Y_{l'}^{m'}(\theta, \phi)} \sin \theta d\theta d\phi = \delta_{l,l'} \delta_{m,m'}$$

where  $\overline{Y_l^m}$  is the complex conjugate of  $Y_l^m$ , and  $\delta$  is the **Kronecker delta**, which equals 1 if the indices are equal, and 0 otherwise.

## 2. Spherical Harmonics Expansion of Functions on the Sphere

Given a scalar field  $f(\theta, \phi)$  defined on the surface of a sphere (such as the Earth's atmosphere), we can expand it in terms of spherical harmonics as a series:

$$f(\theta, \phi) = \sum_{l=0}^L \sum_{m=-l}^l a_l^m Y_l^m(\theta, \phi)$$

Here:

- $L$  is the **truncation level**, which determines the maximum degree of spherical harmonics included in the expansion.
- $a_l^m$  are the **spectral coefficients** that need to be determined.

**Computing the Coefficients:** To find the spectral coefficients, we use the orthonormality of the spherical harmonics. The coefficient  $a_l^m$  is given by the integral:

$$a_l^m = \int_0^\pi \int_0^{2\pi} f(\theta, \phi) \overline{Y_l^m(\theta, \phi)} \sin \theta d\theta d\phi$$

This integral projects the function  $f(\theta, \phi)$  onto the spherical harmonic  $Y_l^m(\theta, \phi)$ , extracting the corresponding coefficient.

**Truncation of the Series:** In practice, we do not include all spherical harmonics. Instead, we truncate the sum at some maximum value of  $l=L$ , which results in a finite series expansion. The truncation determines the **spectral resolution** of the model and defines the number of spherical harmonic modes included in the representation of the field.

For example:

- If  $L = 10$ , then the expansion includes all spherical harmonics from degree  $l = 0$  up to  $l = 10$ .
- The total number of terms in the expansion is given by the sum  $\sum_{l=0}^L (2l+1)$ , which grows roughly as  $L^2$ .

For  $L = 10$ , the total number of terms is:

$$N_{terms} = \sum_{l=0}^{10} (2l+1) = 1 + 3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 + 19 + 21 = 100$$

so, the function  $f(\theta, \phi)$ , is approximated by a sum of 100 spherical harmonic terms.

## 3. Properties of Spherical Harmonics

**Orthogonality and Completeness:** Spherical harmonics are **orthogonal** over the surface of the sphere, meaning that any two different harmonics  $Y_l^m$  and  $Y_{l'}^{m'}$  are perpendicular in the function space. This property makes them ideal for decomposing a function into distinct components.

Additionally, the set of spherical harmonics forms a **complete basis** for any square-integrable function defined on the sphere. This means that any smooth function  $f(\theta, \phi)$  can be represented as a (potentially infinite) sum of spherical harmonics.

**Symmetry:**

- The degree  $l$  of a spherical harmonic determines the **angular scale** of the feature. Large values of  $l$  correspond to smaller-scale features (higher-frequency components), while small values of  $l$  correspond to large-scale, low-frequency components.
- The order  $m$  determines the **longitudinal symmetry**. For each  $l$  there are  $2l + 1$  distinct values of  $m$  ranging from  $-l$  to  $l$  corresponding to different symmetries of the field in the longitudinal direction.

For example:

- The spherical harmonic  $Y_1^0(\theta, \phi)$  corresponds to a dipole, with a variation in latitude but no longitudinal variation.
- The spherical harmonic  $Y_2^1(\theta, \phi)$  corresponds to a pattern with more complex longitudinal symmetry.

**Scaling:** The spherical harmonics have a natural scaling property. For instance, multiplying a spherical harmonic by a constant factor will change the amplitude of the corresponding mode without affecting the symmetry of the function. This property is useful for scaling fields, such as wind speeds or temperature distributions, in climate models.

#### 4. Applications of Spherical Harmonics in Climate Models

In climate and weather models, spherical harmonics are used to represent global fields with high accuracy and relatively few coefficients, which is especially useful for simulating large-scale atmospheric dynamics.

Some specific applications include:

**Atmospheric Dynamics:**

- Atmospheric models often represent variables like pressure, temperature, and wind in terms of spherical harmonics. By expanding these fields in spherical harmonics, models can capture large-scale features such as the **zonal wind**, **jet streams**, and **planetary waves** with fewer coefficients than would be required for grid-based methods.
- For example, the **vorticity equation** or the **primitive equations** of motion are often solved in spectral space, and spherical harmonics are used to expand the relevant fields.

**Global Climate Models (GCMs):**

- In GCMs, the Earth's climate is simulated by solving for variables (e.g., temperature, pressure, wind) at each point on the sphere. By using spherical harmonics, GCMs can efficiently model the global climate while avoiding the need for very fine grids (which would be computationally expensive).
- **Spectral Finite-Difference Models:** Some models combine the benefits of spectral methods for large-scale patterns with finite-difference methods for localized phenomena (like convection or boundary layers), allowing for efficient simulations of both large-scale and small-scale processes.

**Ocean Models:** Similar to atmospheric models, ocean circulation models may use spherical harmonics to represent the spatial structure of the ocean's surface currents, sea level anomalies, and other global patterns.

**Data Assimilation:** In data assimilation methods, spherical harmonics can be used to represent the state of the atmosphere or ocean at a given time, with spectral coefficients used to update the model's state based on observations.

They are eigenfunctions of the laplacian on the field in spectral coordinates, ( $l > m$ , if  $m = 0$ , I have no structure in longitude). Derivatives of spectral harmonics are just linear combinations of other spectral harmonics. NB the sum is from  $-m$  to  $m$  to guarantee the quantity is real and not complex.

$$\nabla^2 \rightarrow \frac{-l(l+1)}{a^2}$$

## A.2 Vector calculus

### A.2.1 Curl

The curl of a vector field measures the tendency of the field to induce rotation around a point, or how much the field "circulates" locally. It is a vector operation applied to a vector field, and the result is another vector field. Mathematically, for a vector field  $\mathbf{F} = (F_x, F_y, F_z)$  the curl is:

$$\nabla \times \mathbf{F} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix}$$

$\hat{i}, \hat{j}, \hat{k}$  are the unit vectors in the  $x, y, z$  directions. The curl results in a vector field whose direction indicates the axis of rotation (using the right-hand rule), and whose magnitude indicates the strength of the rotation. The curl represents the amount and direction of rotational motion or "swirl" in the field. In fluid dynamics, a region of nonzero curl corresponds to vorticity, where fluid particles rotate.

### A.2.2 Gradient

The gradient measures the rate and direction of the steepest increase of a scalar field. It is a vector operation applied to a scalar field, and the result is a vector field. Mathematically, for a scalar field  $f(x, y, z)$ :

$$\nabla f = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$$

The gradient points in the direction where the scalar field  $f$  increases most rapidly.. The magnitude of the gradient gives the rate of increase in that direction. In thermodynamics, the gradient of temperature describes the direction and rate of heat flow. The result of a gradient is always a vector.

- The curl tells you about rotational behavior in a vector field.
- The gradient gives you the direction and rate of steepest ascent in a scalar field.

### A.2.3 Divergence

In vector calculus, divergence measures the rate at which a vector field spreads out or converges at a given point. It quantifies how much the field "diverges" from or "converges" into a region. For a vector field  $\mathbf{F} = (F_x, F_y, F_z)$  the divergence is defined as:

$$\nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

Divergence Positive ( $\nabla \cdot \mathbf{F} > 0$  the field is "spreading out" or "expanding" at a point. Example: A source emitting fluid, such as water spraying out of a fountain. Divergence Negative ( $\nabla \cdot \mathbf{F} <$

0) the field is "converging" or "contracting" at a point. Example: A sink where fluid flows inward, such as water draining into a hole. Zero Divergence ( $\nabla \cdot \mathbf{F} = 0$ ) the field is neither expanding nor contracting, often indicating incompressible flow. Example: Steady, non-divergent flow like incompressible fluid flow.

## A.3 EOF: Empirical Orthogonal Functions

*empirical*= derived from data, *orthogonal*= the scalar product is zero.

**Def.** Tool for analyzing large datasets that vary over both space and time; it helps to extract dominant patterns of variability in a data set.

### How does it work?

1. Matrix representing spatiotemporal data  $X(i, j)$ :

rows  $i \rightarrow$  time steps

columns  $j \rightarrow$  spatial points

2. To identify the major patterns of variability we compute the covariance matrix of the data. It describes how each spatial location's data varies in relation to every other location:

$$C = COV = \frac{1}{n} X^T X$$

where  $X = X(i, j) - \bar{X}_j$

3. The covariance matrix is decomposed into eigenvalues and eigenvectors  $\rightarrow$  SINGULAR VALUE DECOMPOSITION SVD:

$$C\mathbf{v} = \lambda\mathbf{v}$$

with  $\mathbf{v}$  eigenvector or spatial patterns,  $\lambda$  eigenvalues (it describes the variance expressed by each pattern). In this way we get how each data location relates to others.

4. Projecting the original data into these  $\mathbf{v}$  we obtain the PRINCIPAL COMPONENTS that show how each spatial pattern evolves over time.

$$PC(t) = X \cdot \mathbf{v}$$

Therefore,

- EOFs (=eigenvectos) represent the dominant spatial patterns in data, each shows a specific pattern variability over space (e.g. how  $T$  pattern shifts across the globe).
- PCs( principal components) are the time series corresponding to each EOF.

The eigenvalues tell how much variance each EOF captures: if they capture most of the variance it means they represent the major patterns in the data. EOF and PCS (Principal Component Analysis) are the same techniques mathematically but the EOF goal is to find SPATIAL PATTERNS and their correspondig TIME EVOLUTION; while the PCA aims to find the directions of maximum variance in data irrespective of space and time. PCA helps in dimensionality reduction.

**Note.** The diagonal I find after the eigenvalue decomposition is the sum of the covariants:

$$X = U\Sigma V^T \quad \text{so that} \quad S = XX^T = U\Sigma^2 U^T$$

In  $U$  are the eigenfunctions associated to the covariant, in  $\Sigma$  the eigenvalues.

When you have a time serie of data, you should organize them in some order: a matrix where every row ( $1 \leq j \leq n$ ) is a time variation, hence every vector is a space variation of some place in the same tame:

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n]$$

$$\mathbf{S} = \frac{1}{1-n} \mathbf{X} \mathbf{X}^T$$

on the diagonal of  $\mathbf{S}$ , you have the same location (covariance relation), every other element is the covariance of two different locations (it defines how two locations are connected in time):

$$\text{Tr}(\mathbf{S}) = \sum \sigma^2$$

. It becomes a correlation matrix if divided by the standard deviations of the locations.  $\mathbf{X} \mathbf{X}^T$  as symmetric, meaning the eigenvalues are positive and eigenvectors are orthogonal. The eigenvectors represent the EOF. The single value decomposition tells us that any matrix can be written in terms of 2 different other orthogonal matrices:  $\mathbf{X} = U\Sigma V^T$ . I could simply take the data matrix and compute the SVD:

$$x(i) = \sum_i u_i \sigma_i v_i$$

how much covariance the pattern  $u$  of EOF corresponding to the singular value decomposition is responsable for:

$$\mu_i = \frac{\sigma_i^2}{\sum_{i=1}^n \sigma_i^2}$$

### A.3.1 Eigenvalue decomposition → Singular Value Decomposition

It helps to understand what is the weight of the  $u_i$  component in the  $x$  dataset.

orthogonal matrix  $UU^T = I \rightarrow$  they are rotating operating on one vector. You could take any transformation breaking it into 3 processes:

rotation → stretching (diagonal matrix) → rotation

If I can decompose  $\rightarrow S = U\Sigma^2 U^T$ . Note that the  $\text{tr}(\Sigma) = S$ . The number of columns in  $U = \Sigma t_i$  where  $t_i$  is the number of time variations and  $x_J$  the number of features.

Every column in the data matrix can be expanded:

$$x(t) = \sum_i u_i \sigma_i v_i(t)$$

with  $u_i$ = patterns,  $\sigma_i$ =singular value and  $v_i$  weight of the pattern i that particular time.  $x = u_i$  in terms of vector with the same length. EOF method defines only the independent variables.

SST (lat,long)  $\rightarrow$  transforms into a vector  $\begin{bmatrix} lat_1 \\ lat_2 \\ \dots \\ lat_n \end{bmatrix} \rightarrow$  I can represent them in a map.

( $u_i$  and  $v_i$  compensate each other. Higher EOF means more zero lines, meaning more oscillations.

### A.3.2 Limitations

- because of the + and – we are imposing a structure looking for patterns that are orthogonal (we exclude data)
- thinking of EOF as oscillations

### A.3.3 Variance

Variance:

$$\sigma_{VAR} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n} \quad (\text{A.1})$$

### A.3.4 Covariance

Basically it is a straight multiplication. It's a statistical measure that indicates the degree to which two random variables change together. If two variables tend to increase or decrease in tandem then  $COV > 0$ ; if one tends to increase while the other decreases, then  $COV < 0$ .

$$COV(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (\text{A.2})$$

Covariance helps identify the relationship between two variables but it does not provide information about the strength or scale of the relationship → correlation.

### A.3.5 Correlation

Statistical measure that describes the strength and direction of a linear relationship between two variables. It STANDARDIZES the relationship, providing a dimensionless value that ranges between –1 and 1.

$$r = \frac{COV(X, Y)}{\sigma_x \sigma_y} \quad (\text{A.3})$$

$\sigma_i$  are the standard deviations.

- $r = 1$  perfect \_positive linear relation
- $r = -1$  perfect \_negative linear relation
- $r = 0$  no linear relation

## A.4 Space-time splitting

The dominant shape of the global circulation suggests that some understanding can be gained from splitting the physical fields into larger and smaller portions using appropriate averages. At a first inspection, the flow is seen as a predominant circumpolar vortex with superposed fluctuations in space and time. The longitudinal direction is also known as the “zonal” direction, therefore the average over longitude is known as the zonal averaging

### A.4.1 Zonal mean

The zonal mean of a quantity A is defined as the average over longitudes. Commonly used symbols in the literature are the overbar  $\bar{u}$  or square brackets  $[u]$ , the first is most frequently

encountered in the theoretical and modeling literature whereas the second is most commonly used in observational and diagnostics works. In the following, we will use brackets.

$$[A] = \frac{1}{2\pi} \int_0^{2\pi} A dx \quad (\text{A.4})$$

so that the entire field can be decomposed in

$$A = [A] + A^* \quad (\text{A.5})$$

where  $A^*$  is the deviation from the zonal mean. The average has the properties that  $[[A]] = [A]$  and  $[A^*] = 0$ .

When a stream function can be defined, the average zonal mean meridional velocity is zero:

$$= \frac{1}{2\pi} \int_0^{2\pi} v dx = \frac{1}{2\pi} \int_0^{2\pi} \frac{\partial \psi}{\partial x} dx = 0 \quad (\text{A.6})$$

This is a consequence of the more general result that the zonal mean of any quantity that is a longitude derivative is zero.

#### A.4.2 Time mean

The time mean is defined simply as the average over a length of time. Also in this case several common symbols are used, once again the overbar or sometimes curly brackets, in the following we will use the overbar

$$\bar{A} = \frac{1}{T} \int_0^T A dt \quad (\text{A.7})$$

so that the total field is

$$A = \bar{A} + A' \quad (\text{A.8})$$

The deviations from the zonal means are called "eddy" components. An eddy that obeys a dispersion relation is a "wave".

#### A.4.3 Higher order quantities

The averages can be used to decompose higher-order quantities. For instance using zonal means a quadratic correlation of the form  $AB$  can be decomposed as

$$AB = ([A] + A^*)([B] + B^*) = A^*B^* + [A]B^* + A^*[B] + [A][B] \quad (\text{A.9})$$

taking a further zonal mean we get

$$= [A^*B^* + [A]B^* + A^*[B] + [A][B]] = [A][B] + [A^*B^*] \quad (\text{A.10})$$

because the mix terms disappear as the average of the deviations is zero.

We can refine the splitting by considering the time average splitting of the zonal terms:

$$= [\bar{A}] + [A]'A^* = \bar{A}^* + A'^* \quad (\text{A.11})$$

These terms represent the stationary symmetric circulation, the transient symmetric circulation and the stationary deviation from the zonal means ("asymmetries") and the transient asymmetries.

Inserting these relations into Eq. A.10 we get

$$= ([\bar{A}] + [A]')([\bar{B}] + [B]') + [A^*B^*] \quad (\text{A.12})$$

the time mean of the terms linear in the time deviation will average again to zero (this time with respect the time mean) and we finally get

$$\overline{[AB]} = [\bar{A}][\bar{B}] + \overline{[A]'}[\overline{B}'] + \overline{[A^*B^*]} \quad (\text{A.13})$$

The decompositions are not unique. We have first performed the split in the zonal mean and then the split in the time mean, considering a split only in the eddy part:

$$A^* = \bar{A}^* + A'^* \quad (\text{A.14})$$

we would get

$$\overline{AB} = [\bar{A}][\bar{B}] + [\bar{A}^*\bar{B}^*] + \overline{[A'^*\bar{B}^*]} \quad (\text{A.15})$$

where the first term is the contribution of the mean meridional circulation, the second is the contribution of the time-mean (standing) eddies and the last one is the contribution of the transient eddies. This kind of decomposition is therefore a useful instrument but requires always consideration of the hypothesis formulated in the initial design. Another consideration is that they depend on the specific kind of averaging that is used. There is little choice in the zonal mean, being fixed by the geometry, but we have much more choices in the case of the time mean. Results will depend on the length of the time averaging period and on the original frequency of the data. Time mean and second-order quantities calculated over daily will differ from the same quantities calculated over time series of weekly or monthly data. There is no *correct* choice, each one will offer a different glimpse in the data from a chosen perspective.



# Bibliography

- B.J.Hoskins, D.J.Karoly (1981). "The Steady Linear Response of a Spherical Atmosphere to Thermal and Orographic Forcing". In: *Journal of the Atmospheric Sciences*.
- Grose, W.L. and B.J. Hoskins (1978). "On the Influence of Orography on Large-Scale Atmospheric Flow". In: *Journal of the Atmospheric Sciences*.
- I.Held, A.Hou (1980). "Nonlinear Axially Symmetric Circulation in a Nearly Inviscid Atmosphere". In: *Journal of the Atmospheric Sciences*.
- Manabe, Syukuro (1964). "Thermal Equilibrium of the Atmosphere with a Convective Adjustment". In: *Journal of the Atmospheric Sciences*.
- Manabe, Syukuro (1967). "Thermal Equilibrium of the Atmosphere with a given Distribution of Relative Humidity". In: *Journal of the Atmospheric Sciences*.
- Manabe, Syukuro and K. Bryan (1969). "Climate Calculations with a Combined Ocean-Atmosphere Model". In: *Journal of the Atmospheric Sciences*.
- Manabe, Syukuro and R.T. Wetherald (1975). "The Effects of Doubling the CO<sub>2</sub> Concentration on the Climate of a General Circulation Model". In: *Journal of the Atmospheric Sciences*.
- Richardson, L.F. (1922). *Weather Prediction by Numerical Processes*. Cambridge University Press.
- Rowell, David P. (1995). "Assessing Potential Seasonal Predictability with an Ensemble of Multidecadal GCM Simulations". In: *Journal of the Atmospheric Sciences*.
- Stone, D.A. and J. Weaver (2000). "Projection of Climate Change onto Modes of Atmospheric Variability". In: *Journal of the Atmospheric Sciences*.
- Tanner, P. Cvitanović R. Artuso R. Mainieri G. and G. Vattay (2020). *Chaos: Classical and Quantum*. Niels Bohr Institute, Copenhagen.



