Mathematical Modeling for Climate and Environment
SS 2020

Prof. Dr. Vadym Aizinger

December 22, 2020

Contents

Literature

- H. Goosse, *Climate System Dynamics and Modelling*, Cambridge University Press, 2015
- H. Goosse et al., *Introduction to climate dynamics and climate modelling*, www.climate.be/textbook
- J. B. Drake, *Climate Modeling for Scientists and Engineers*, SIAM, 2014
- T. Slawig, *Klimamodelle und Klimasimulationen*, Springer, 2015
- H. Kaper, H. Engler, *Mathematics & Climate*, SIAM, 2013
- T. Stocker, *Introduction to Climate Modelling*, Springer, 2011
- B. Cushman-Roisin, J.-M. Beckers, *Introduction to geophysical fluid dynamics*, Elsevier, 2011
- M. Jacobson, *Fundamentals of Atmospheric Modeling*, Cambridge University Press, 2005

Acknowledgments
Huge thanks to Hugues Goosse for a great textbook and personal support, to Alexander Prechtel for slides on subsurface modeling, and to Ian Hewitt for a nice intro into Continuum Mechanics.

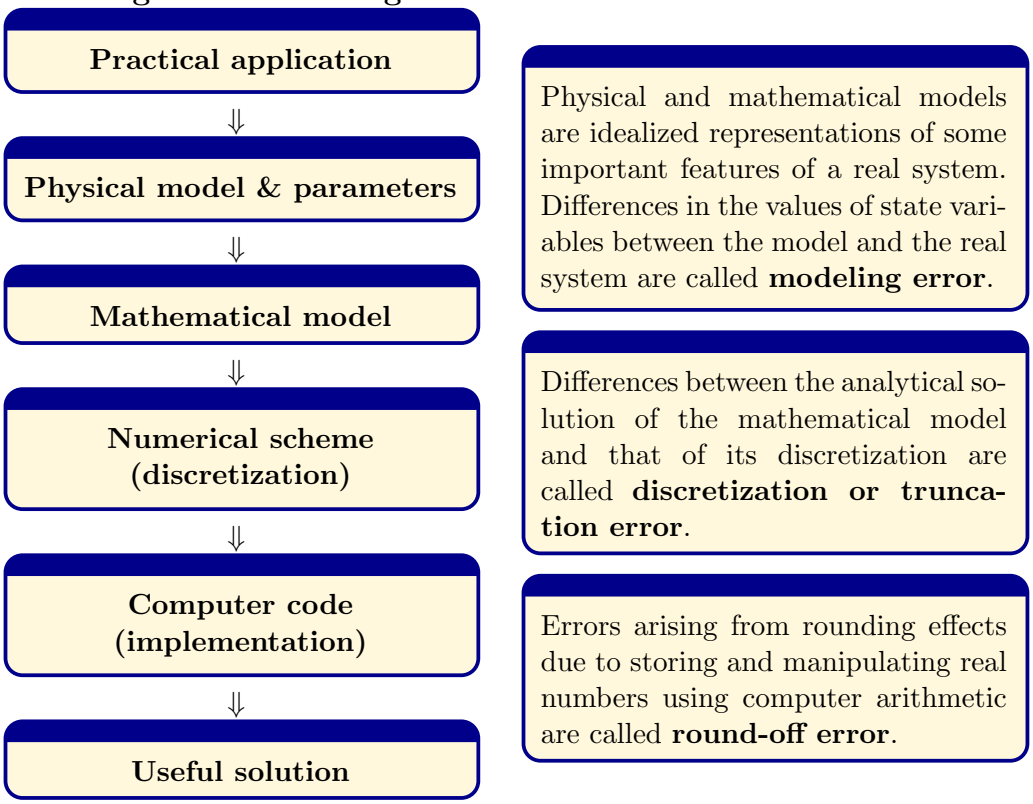
1 Fundamentals

1.1 Principles of mathematical modeling

Concept of a mathematical model

Definition 1.1. *A mathematical model is a description of a system using mathematical concepts and language. This description is given by specifying a set of* **governing equations** *supplemented by appropriate* **boundary and initial conditions**.
The governing equations express the relationships between the **state variables** *of the system (i.e. those that completely describe the mathematical 'state' of the model) and may also depend on empirical* **parameters**, *random variables, or other inputs.*

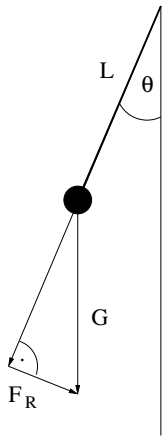
Modeling: Workflow diagram and error sources



Modeling: Example application

Computation of movement of a pendulum

- L: Length of the pendulum arm
- m: Mass of the pendulum
- g: Acceleration of free fall
- $G = mg$: Gravity force
- θ: Angle of deviation
- $F_R = G \sin\theta$: Restoring force



Modeling: Physical model

Assumptions

- Mass point instead of pendulum head
- Pendulum arm – zero mass, zero deformation
- No air resistance (vacuum) and no friction in the pendulum arm

Physical laws and parameters

- Gravitational acceleration at Earth's surface $g \approx 9.8\ [m/s^2]$
- Angular velocity $\omega(t) = \theta'(t)$
- Velocity of the pendulum is given by $v(t) = L\omega(t)$
- Acceleration of the pendulum is given by $v'(t) = L\omega'(t) = L\theta''(t)$
- Restoring force $F_R(t) = -mg \sin\theta(t)$
- On the other hand: $F_R(t) = mv'(t) = mL\theta''(t)$ (Newton's 2nd law)

Modeling: Mathematical model

Putting everything together gives an **ordinary differential equation**

(ODE)

$$\theta''(t) = -\frac{g}{L}\sin\theta(t)$$

that also needs initial conditions:

$$\theta(0) = \theta_0 \quad \theta'(0) = \omega(0) = \omega_0\,.$$

Remark 1.2. *This differential equation does not have a solution $\theta(t)$ expressible in terms of elementary functions (which is the case for many ODEs). Therefore this equation can only be solved numerically.*

Modeling: Discretization

- Approximate $\theta'(t)$ and $\omega'(t)$ by differential quotients (with a small $h > 0$)

$$\theta'(t) \approx \frac{\theta(t+h)-\theta(t)}{h}\,, \quad \omega'(t) \approx \frac{\omega(t+h)-\omega(t)}{h}\,.$$

- On the other hand, we know that $\quad \theta'(t) = \omega(t)\,, \quad \omega'(t) = -\frac{g}{L}\sin\theta(t)\,.$
- Now, using the **explicit Euler method** we can compute from $(\theta(t), \omega(t))$ an approximation to $(\theta(t+h), \omega(t+h))$:

$$\begin{aligned}\theta(t+h) &\approx \theta(t) + h\omega(t)\\ \omega(t+h) &\approx \omega(t) + h\left(-\frac{g}{L}\sin\theta(t)\right)\end{aligned}$$

Modeling: Computer program

```
function pendel
theta0 = 0; omega0 = 6.23; T = 10; N = 10000;
[theta, time, omega] = stabpendel(theta0, omega0, T, N);
plot(time,theta);
end

function [theta, time, omega] = stabpendel(theta0, omega0, T, N)
g=9.81; L=1.0; h = T/N;
time = zeros(N, 1); theta = zeros(N, 1); omega = zeros(N, 1);
```

```
theta(1) = theta0; omega(1) = omega0;

for i = 1 : N
    time(i+1) = time(i)+h;
    theta(i+1) = theta(i)+h*omega(i);
    omega(i+1) = omega(i)-h*g/L*sin(theta(i));
endif
end
```

Modeling: Verification and validation

Definition 1.3. *Verification is the process of establishing the correctness of the discretization and the computer code of a numerical model. The main criterion is the ability of the implementation to accurately approximate the mathematical model being discretized.*

- The parts affected by the verification procedure include the numerical scheme and the computer code
- Verification procedures frequently rely on problems with known solutions, convergence studies, various software checking tools, etc.
- For complex non-linear models, the **method of manufactured solution** is often very useful

Definition 1.4. *Validation examines the fitness of the model with respect to the intended purpose: Faithfully imitating some important features of the real system.*

- The validation process can affect all aspects of the model development workflow:
  - Physical models and assumptions behind them
  - mathematical models including equations, boundary conditions, parametrizations, etc.

  - choice of numerical method, order of the approximation scheme, and computational mesh
  - the model code including the computational performance aspects

- Validation techniques are **benchmarks** of various degrees of realism and comparisons to the results of established models and observations
- A critical criterion of model's validity is its ability to make **predictions**

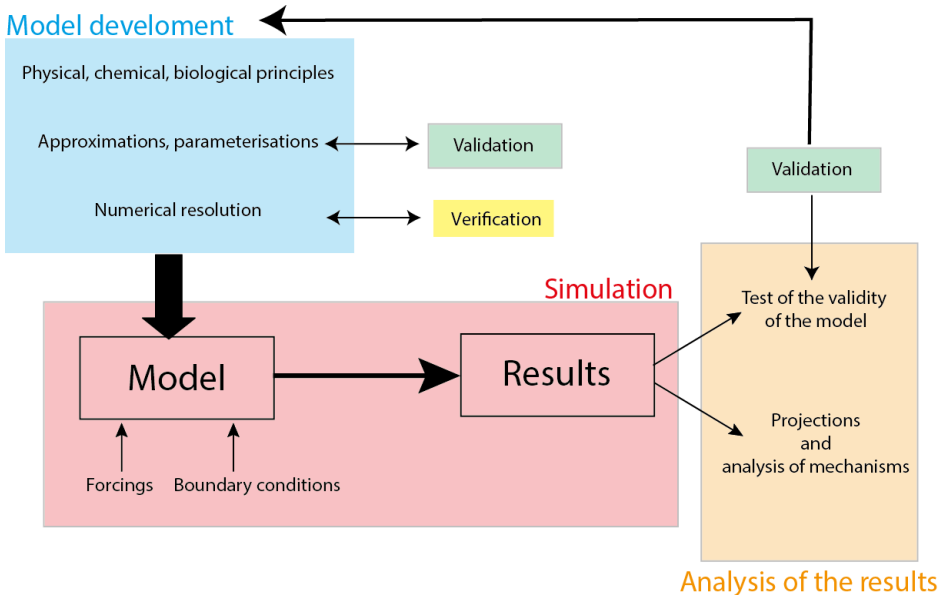
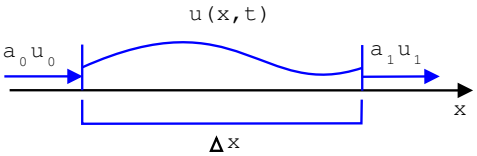


Figure 1.1: Schematic of verification vs. validation (Goosse 2015).

Modeling: Conservation law in one dimension



$u(x,t)$ – concentration of dissolved substance, $a(x,t)$ – the flow velocity.

Consider the conservation of mass of $u(x,t)$ in control volume Δx:

$$\int_{\Delta x} u(x,t+\Delta t)dx = \int_{\Delta x} u(x,t)dx + \int_t^{t+\Delta t} \left(a_0(\tau)u_0(\tau) - a_1(\tau)u_1(\tau)\right) d\tau$$

$$\Downarrow /(\Delta x\,\Delta t)$$

$$\frac{1}{\Delta x}\int_{\Delta x}\frac{u(x,t+\Delta t)-u(x,t)}{\Delta t}dx=-\frac{1}{\Delta t}\int_{t}^{t+\Delta t}\frac{a_1(\tau)u_1(\tau)-a_0(\tau)u_0(\tau)}{\Delta x}d\,\tau$$

$$\Downarrow \Delta x\to 0,\ \Delta t\to 0$$

$$\frac{\partial u}{\partial t}=-\frac{\partial(a\,u)}{\partial x}\quad\Leftrightarrow\frac{\partial u}{\partial t}+\frac{\partial(a\,u)}{\partial x}=0$$

Modeling: Conservation law in multiple dimensions

In the multidimensional case, the **integral form** of conservation law holds:

$$\partial_t\int_{\widetilde{\Omega}}u(x,t)\,dx=-\int_{\partial\widetilde{\Omega}}\boldsymbol{a}(x,t)\cdot\boldsymbol{n}(x)\,u(x,t)\,d\sigma+\int_{\widetilde{\Omega}}Q(x,t,u(x,t))dx.\quad(1.1)$$

$\boldsymbol{n}$ is the exterior unit normal to $\partial\widetilde{\Omega}$, Q is sources/sinks of u.
From the Gauss theorem, one has

$$\int_{\widetilde{\Omega}}\{\partial_t u(x,t)+\nabla\cdot(\boldsymbol{a}(x,t)\,u(x,t))-Q(x,t,u(x,t))\}\,dx=0.$$

Given sufficient regularity and requiring the above to hold for arbitrary $\widetilde{\Omega}\subset\Omega$ results in the **differential form** of the conservation law:

$$\partial_t u(x,t)+\nabla\cdot(\boldsymbol{a}(x,t)\,u(x,t))=Q(x,t,u(x,t)),\ x\in\Omega,\ t\in(0,T].\quad(1.2)$$

Modeling: Diffusion equation

Remark 1.5. *Conservation laws describe the process of transport (advection, convection). Another common physical process taking place in fluids is* **molecular diffusion** *arising from Brownian motion.*

Consider now a vessel with fluid at rest. For, $J^{(1)}$ the **diffusive flux** of u, the empirical measurements indicate (Fick's law)

$$\boldsymbol{J}^{(1)}=-\boldsymbol{K}\nabla u,\quad(1.3)$$

where $\boldsymbol{K}$ [m^2/s] is the positive definite **molecular diffusivity** tensor.
The resulting relationship is called the **diffusion equation**

$$\partial_t u - \nabla \cdot (\boldsymbol{K} \nabla u) = Q. \tag{1.4}$$

The same type of equation is also used to model heat transport in solids.

Modeling: Advection-diffusion equation
If the fluid is not at rest but is stirred with given velocity $\boldsymbol{a}$ we have, in addition, an **advective (convective)** flux described by

$$\boldsymbol{J}^{(2)} = \boldsymbol{a} u. \tag{1.5}$$

Together, one obtains an **advection-diffusion** equation:

$$\partial_t u + \nabla \cdot (\boldsymbol{a} u - \boldsymbol{K} \nabla u) = Q. \tag{1.6}$$

For stationary processes, (1.6) becomes

$$\nabla \cdot (\boldsymbol{a} u - \boldsymbol{K} \nabla u) = Q$$

and, for scalar constant $K = 1$, $\boldsymbol{a} = 0$ and $f := Q$, independent of u, we get the **Poisson equation**

$$-\Delta u = f \quad \text{in} \quad \Omega\,.$$

Modeling: Nonlinear advection-diffusion equation
In a more general form, advection, diffusion, and source/sink terms can nonlinearly depend on u (in addition to being functions of time and spatial coordinates)

$$\partial_t u + \nabla \cdot (\boldsymbol{A}(t, x, u) - \boldsymbol{K}(t, x, u, \nabla u)) = Q(t, x, u). \tag{1.7}$$

Modeling: Initial and boundary conditions

To have a unique solution, time-dependent problems need initial conditions

$$u(x,0) = u_0(x) \quad x \in \Omega. \tag{1.8}$$

Both, stationary and time-dependent problems need boundary conditions. The standard boundary condition types on $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N \cup \partial\Omega_3$ are

- Dirichlet – specifies the value of u

$$u = u_D, \quad x \in \partial\Omega_D, \tag{1.9}$$

- Neumann – specifies the normal flux of u

$$(\boldsymbol{A}(u) - \boldsymbol{K}(\nabla u)) \cdot \boldsymbol{n} = g_N, \quad x \in \partial\Omega_N, \tag{1.10}$$

- mixed (also known as Robin or Poincare) – specifies a linear combination of u and its normal flux

$$-(\boldsymbol{A}(u) - \boldsymbol{K}(\nabla u)) \cdot \boldsymbol{n} + \alpha u = g_3, \quad x \in \partial\Omega_3. \tag{1.11}$$

Modeling: Analytical solution techniques

Remark 1.6. *There exist* **analytical** *methods for PDEs.* **Classical solutions** *obtained by these methods satisfy equations* **pointwise** *and posses certain* **minimum regularity** *(e.g., $u \in C^2(\Omega)$ for the advection-diffusion equation).*

The best-known methods are:

- Separation of unknowns followed by development into Fourier series
- Taylor series approaches

These methods work for simple (mostly linear with constant coefficients) PDEs on standard domains (rectangles, spheres, etc.).

Remark 1.7. *Nearly all problems of interest in climate (and more generally fluid dynamics) modeling can only be solved* **numerically***!*

Modeling: Numerical methods for PDEs

Idea: Approximate the solution to the PDE that lies in some (as a rule **infinite-dimensional**) function space by a series of **discrete** solutions that belong to **finite-dimensional** function spaces and **converge** to the classical solution in some norm as these discrete spaces are enlarged.

- In this context, two critical points require particular care:
  - Criteria for the quality of approximation (usually in the form of a **norm**)
  - Choice of finite-dimensional function spaces
- Each large class of numerical methods for PDEs is characterized by a specific combination of the above two points
- Another important decision factor is the **computational efficiency** of the chosen numerical scheme

Modeling: Approximation quality – choice of the norm

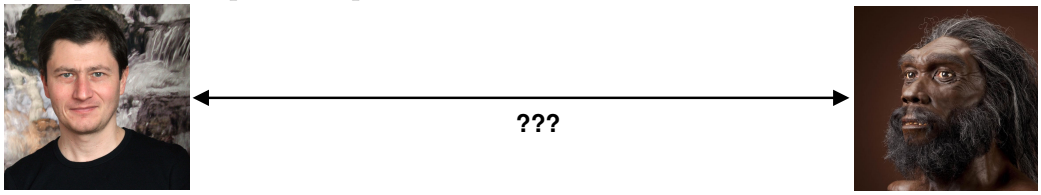
Evaluating the quality of the approximation is a critical issue!

When comparing the analytical and numerical solutions one has

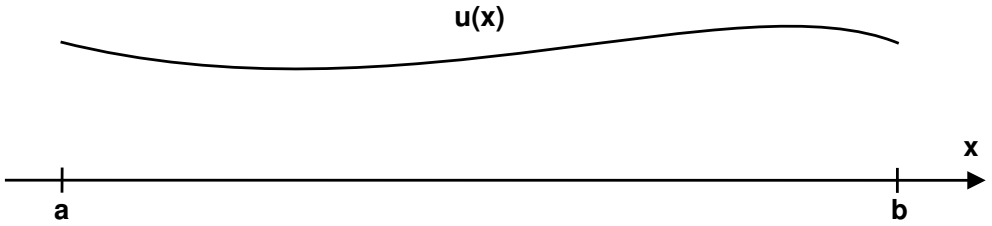
- two functions
- that lie (as a rule) in **different function spaces**

A simple example:

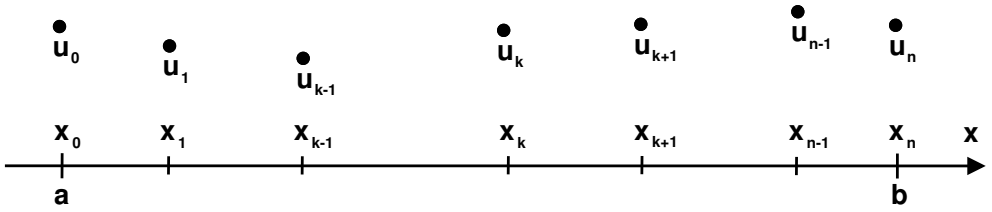
Not quite as simple example:



Modeling: Approximation space – finite differences
Given a smooth function $u(x)$



the finite difference method approximates the values of u in selected **nodes**



Main features of finite difference methods

- Discretizes the PDE directly (in differential form)
- Standard error norms: l^∞ and l^1 (discrete norms for vectors in $\mathbb{R}^n$)

\+ Simple to implement and analyze

\+ Can produce computationally efficient and easily vectorizable code

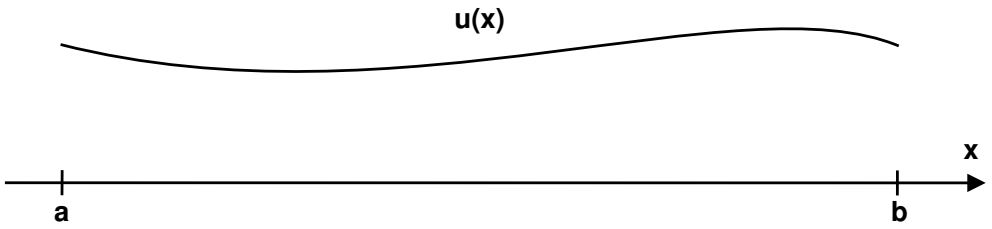
\+ High-order approximations possible (mostly by extending the stencil)

– In multidimensions generally restricted to structured grids or those that are topologically equivalent to structured

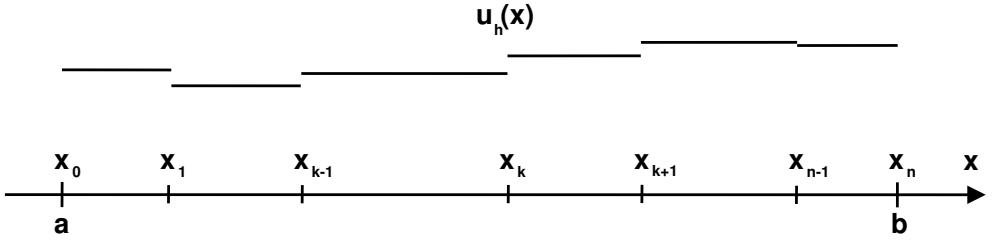
– High-order discretizations need wider stencils impairing parallel scaling

– Solution is only available in grid nodes – evaluations at all other points require **interpolation**

Modeling: Approximation space – finite volumes



For a given partition of the computational domain into **cells**, the finite volume method approximates the mean value of the solution on each cell[1].



Main features of finite volume methods

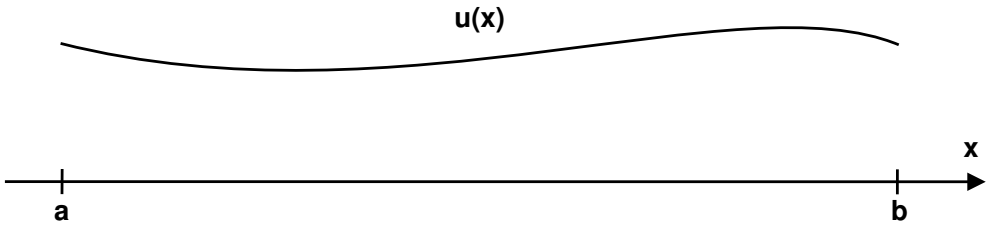
- Discretizes the PDE in integral form (after integration by parts)
- Standard error norm: $L^1(\Omega_h)$

+ Highly suitable for unstructured meshes
+ Robust and stable for advection-dominated problems
+ Guarantee the local conservation of primary unknowns
+ Can be implemented in a computationally efficient and vectorizable fashion

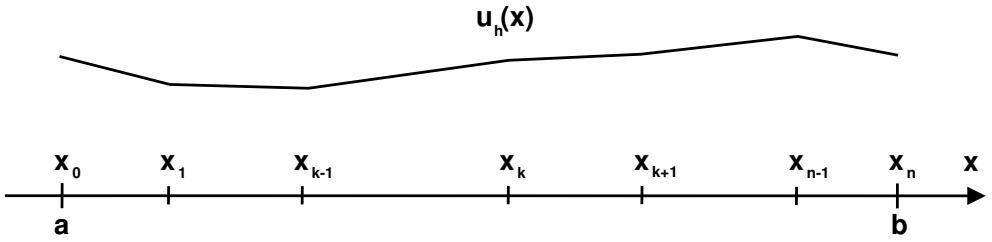
– Difficult to analyze
– High-order discretizations require **reconstruction techniques** with wider stencils that
  – are complicated to implement for unstructured meshes
  – can affect parallel scaling

[1]The cell-centered finite volumes are implied here.

Modeling: Approximation space – (classical) finite elements



For a given partition (triangulation) of the computational domain into **elements**, the finite element method uses **globally continuous**, piecewise polynomials of order $p \geq 1$ defined on this partition as approximation.



Main features of classical finite element methods

- Discretize the PDE in weak (variational) form
- Standard error norm: $L^2(\Omega_h)$

\+ Excellent analysis framework exists based on Sobolev space theory

\+ Highly suitable for unstructured meshes

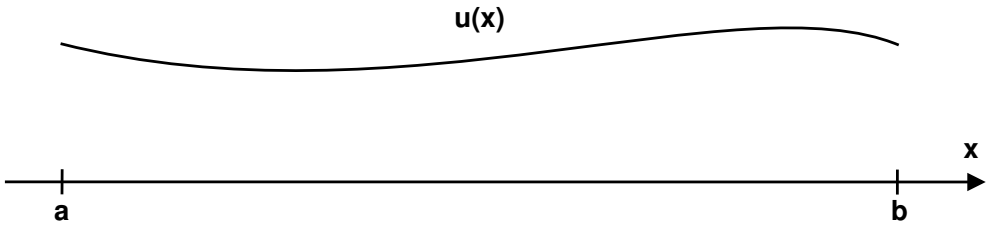
± Computational efficiency is somewhere between that of the finite volume and discontinuous Galerkin methods

± High-order discretizations are possible but not trivial to stabilize and implement

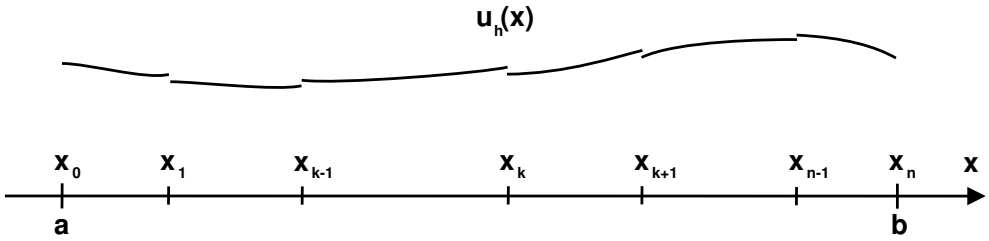
– Needs artificial stabilization for advection-dominated problems

– Does not guarantee the local conservation of primary unknowns

Modeling: Approximation space – discontinuous Galerkin



The discontinuous Galerkin method works similarly to the classical finite elements but uses **globally discontinuous** piecewise polynomial approximation spaces.



Main features of discontinuous Galerkin finite element methods

- Discretize the PDE in weak (variational) form
- Standard error norm: $L^2(\Omega_h)$

\+ Use the same analysis framework as the classical finite elements

\+ Highly suitable for unstructured meshes, high-order discretizations, and all types of adaptivity

\+ Very robust for advection-dominated problems, the local conservation of all primary unknowns is guaranteed

– Much more computationally expensive than the classical finite elements of the same order

± Computational expenses can be somewhat offset by excellent parallel scaling properties

Questions

- What is the purpose of the verification procedure? What standard verification steps/tools do you know?
- Why is the goal of model validation, and what model parts does it concern?
- What are the advantages/disadvantages of using analytical/numerical solution methods for PDEs?
- What are conservation laws and why are they critically important for climate modeling?
- Name the main selection criteria for choosing a numerical method for a PDE-based model.
- Discuss pros and cons for the following discretization methods:
  - finite differences
  - finite volumes
  - finite elements
  - discontinuous Galerkin

1.2 Climate system

Weather and climate

Definition 1.8.
- *Traditionally, climate is defined as a description in terms of* **mean and variability** *of the relevant atmospheric variables: Temperature, atmospheric pressure and humidity, precipitation, cloud cover, wind*
- *According to the World Meteorological Organisation (WMO),* **30 years** *is the minimum required period for performing the statistics used to define climate; however, this period is only indicative*
- *In a more general sense, climate is nowadays frequently defined as the description of the whole Earth climate system*

Remark 1.9. *This contrasts the first interpretation of climate to that of weather – the latter is usually understood as a snapshot of temperature, pressure, etc.*

Climate system: Compartments

- Ocean
- Atmosphere
- Sea ice
- Ice sheets
- Land surface
- Biosphere (consisting of marine and terrestrial parts)

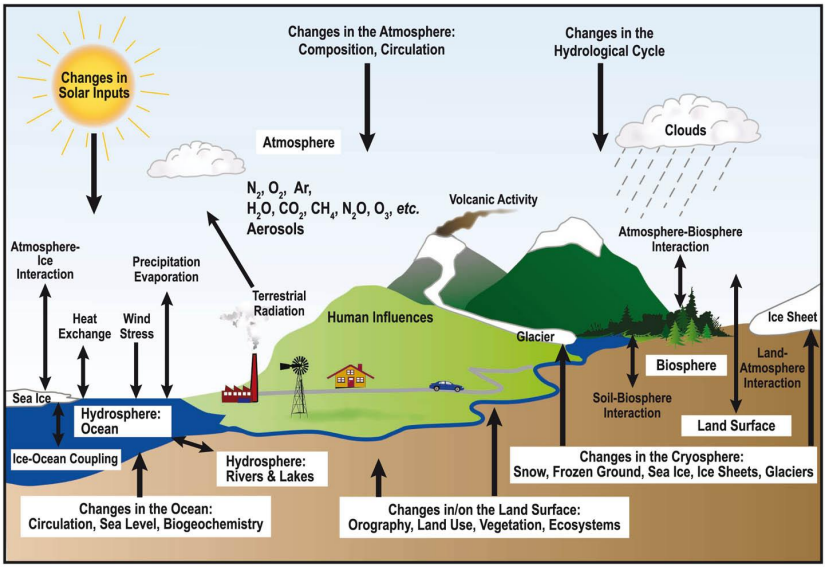


Figure 1.2: Components of the climate system and their interactions (IPCC 2007).

Climate system: Driving forces

Earth climate is a dynamic (temporally varying) system driven by the entering energy (mostly solar) and exchanges of mass, momentum, and energy between its components.

The dynamics of the system is strongly influenced by the following factors:

- total energy budget and its spatial and temporal distribution
- astronomical factors (Earth orbit and rotation, Sun and Moon gravity)
- hydrological (water) cycle
- cycles of major greenhouse gases (carbon dioxide, methane) and aerosols
- geothermal forcings (volcanoes, seismic activity)
- geochemical, biogeochemical, biogeophysical, and antropogenic processes

Climate system: Feedbacks

Earth climate is a system in **dynamic equilibrium**; it responds to **perturbations** by adjusting its equilibrium state.

Definition 1.10. *A special type of system response to a perturbation that affects the perturbing factor itself is called a* **feedback**.

Feedbacks are quantifiable and generally subdivided into two large groups:

- positive feedbacks – i.e. those that amplify the response of the system to the initial perturbation
- negative feedbacks – i.e. those that lower the response of the system to the initial perturbation

The main feedbacks considered in the climate research are those related to the amount and distribution of energy. The response of the climate system as a whole to perturbations in the energy budget must be characterized by a **negative** feedback, otherwise the system would be **unstable**.

Ice-albedo feedback

Example 1.11. *Increase in temperature ⇒ increase in ice melt ⇒ decrease in ice cover / albedo ⇒ increase in absorbed solar radiation ⇒ increase in temperature*

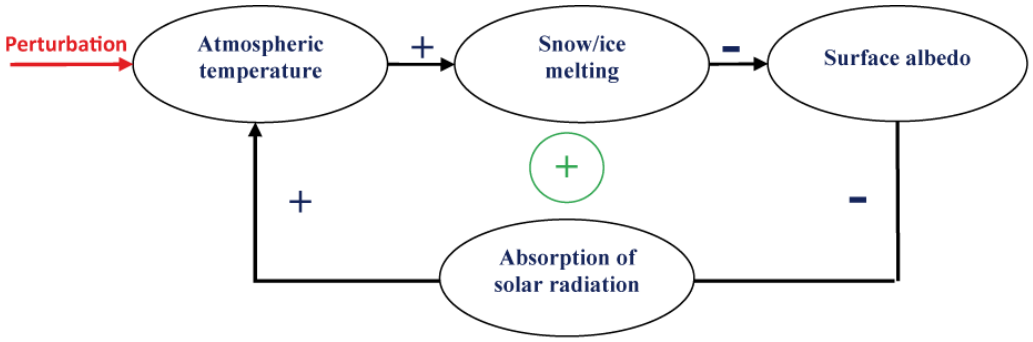


Figure 1.3: Schematic of ice-albedo feedback (Goosse 2015). +/- signs denote the same-direction and the opposite-direction dependencies, respectively.

1.3 Climate models

Climate models: Classification

Definition 1.12. *A climate model is defined as a mathematical representation of the climate system based on physical, biological, and chemical principles.*

Climate models are classified using one or more of the following criteria:

- compartments included (ocean, ocean + atmosphere, etc.)
- modeling focus (process study, general climate)
- spatial coverage (regional or global)
- temporal coverage (present-day climate, paleoclimate, projection, ...)
- number of spatial dimensions (zero-, one-, two-, or three-dimensional)
- stationary or time-varying

Climate models: Hierarchy

In the order of decreasing complexity, the climate models roughly fall into one the following four classes:

- general circulation models (GCM, three-dimensional)
- Earth system Models of Intermediate Complexity (EMIC, usually two- or three-dimensional)
- box models (use boxes for compartments of climate system or large parts of a compartment, e.g. an ocean or a continent)
- energy balance models (EBM, zero- or one-dimensional)

Climate models: General Circulation Models (GCM)

Idea: Use three-dimensional continuum-based representations of the ocean and atmosphere enhanced by any number of additional compartments.

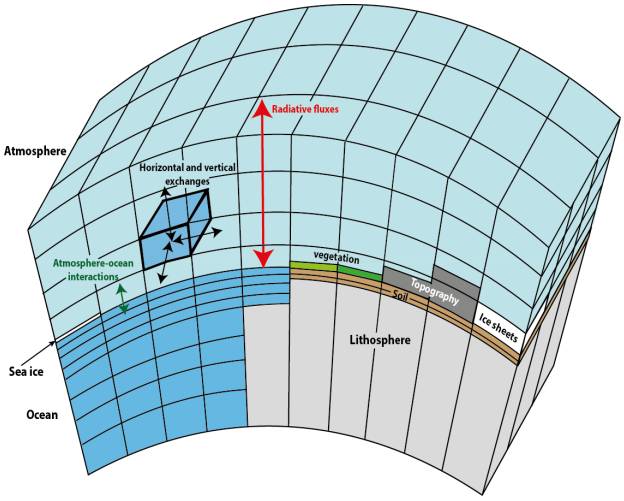


Figure 1.4: Schematic of a General Circulation Model (Goosse 2015).

Climate models: Specific features of GCMs

- Very large spatial extent and often long simulation times
- Large number of physical, chemical, and biological processes of significance that involve time scales between fractions of a second (e.g., some chemical reactions) and millennia (e.g., sedimentation)
- Limited availability of key data, experimental model validation as a rule not possible

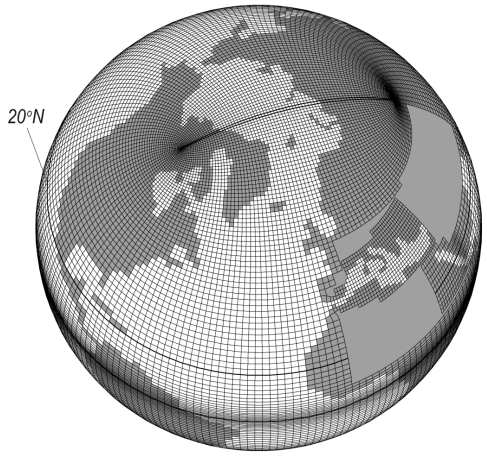


Figure 1.5: Structured mesh from NEMO ocean model (2008).

- Computationally very expensive; in many cases, resolution sufficient for error control/convergence is not achievable in practice

Climate models: Components and naming conventions
GCMs consist of more components than their counterparts from other application areas. The main components have the following names:

- **Dynamical core** is the main part of the model that includes the discretization and the solution algorithm for the PDE system concerned with the key properties of the system (usually, conservation of mass, momentum, energy, etc.)
- Auxiliary models for significant but not explicitly resolved processes are called **parametrizations**. These sub-models are often simpler than the dynamical core and can use empirical or statistical methods
- The model inputs can come in form of
  - **forcings** – usually time-varying user-specified functions utilizing measured data or results of other models
  - **boundary conditions** – data prescribed at the external domain boundaries in form of fluxes or values of primary variables

Climate models: EMIC modeling approaches

Idea: Combine highly simplified approximations for some processes or compartments with sophisticated representations of processes or compartments of interest.

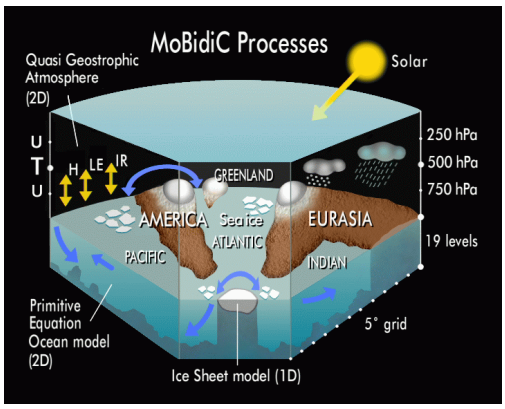


Figure 1.6: Schematic of MOBIDIC model which uses zonally averaged vertical slice models for ocean and atmosphere (Crucifix et al. 2002).

Climate models: Box model of the North Atlantic Ocean

Idea (Stommel, 1961): Represent the southern and the northern parts of the North Atlantic Ocean by separate well-mixed boxes connected by a channel (surface currents) and a pipe (bottom currents).

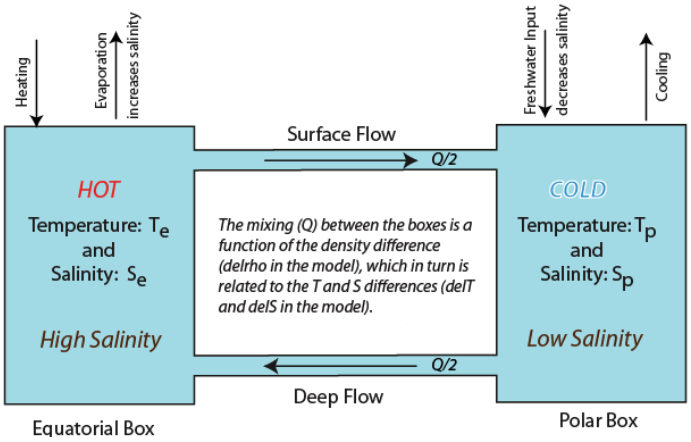


Figure 1.7: Stommel two-box model of North Atlantic circulation (D. Bice).

Climate models: Energy balance models

Changes in heat storage (T) = absorbed solar radiation - emitted terrestrial radiation + net horizontal transport

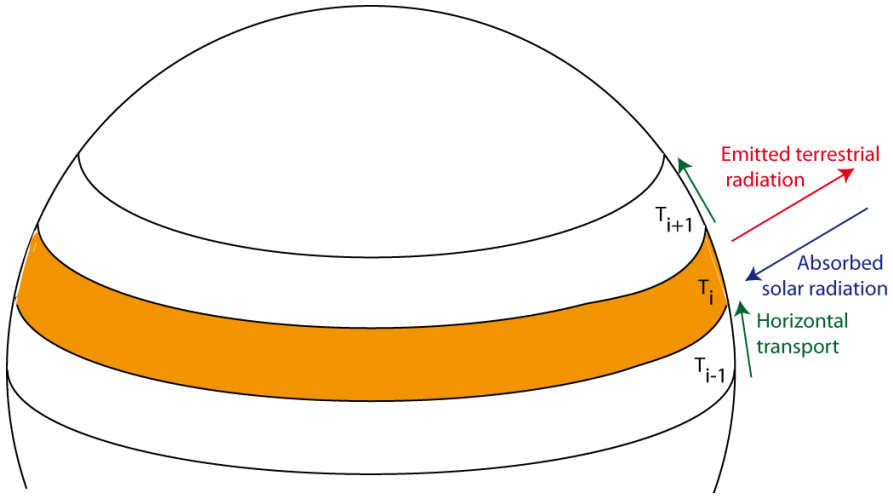


Figure 1.8: A latitude-averaged one-dimensional EBM (Goosse 2015).

Climate models: Regional models

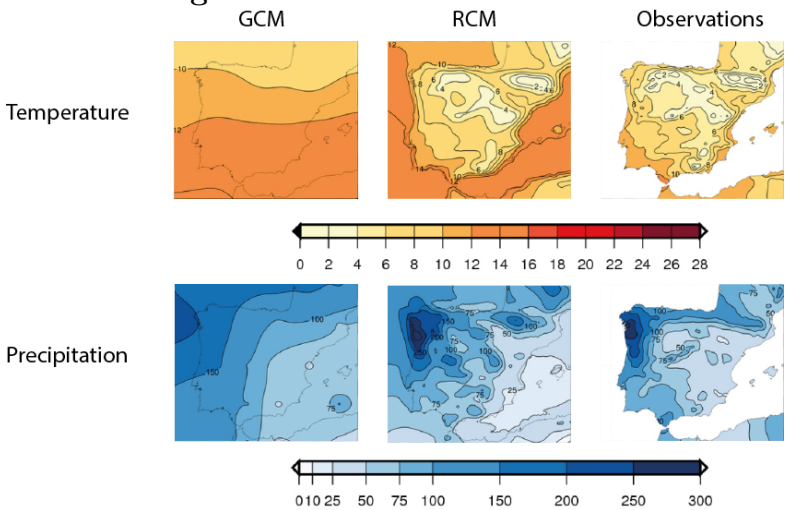


Figure 1.9: Winter temperature [°C] and precipitation [mm/month] in Spain. GCM (resolution ≈400 km) vs. regional climate model (RCM, resolution ≈30 km) vs. observations (Gómez-Navarro et al., 2011).

Questions

- What is climate and how is it related to weather?
- What is the minimum period required to compile the climate statistics?
- Which is the most important compartment of the climate system and why?
- Name the main cycles in the climate system.
- What factors determine the dynamics of the climate system?
- What is a feedback and how can it affect the equilibrium of the climate system?
- Is the Earth climate a stable or an unstable dynamic system and why?
- Name common classification criteria for climate models?
- Discuss the motivation behind the climate model hierarchy.
- What is the dynamical core of a general circulation model and how does it differ from a parametrization?

1.4 Incorporating observations into models

Model evaluation: Performance metrics

The main purpose of **performance metrics** is to provide a quantitative criterion for the quality of model's results. Simplest metrics are based on comparisons between simulation results and observations in some norm. In this context, the l^2 norm in $\mathbb{R}^n$ (also called root-mean-squares, RMS)

$$Err(\boldsymbol{T}_{sim}, \boldsymbol{T}_{obs}) := \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(T_{sim,k} - T_{obs,k}\right)^2}$$

is particularly widely used. Here n denotes the number of observations, and $\boldsymbol{T}_{sim}, \boldsymbol{T}_{obs}$ are the vectors of simulated and observed values, respectively.

Model evaluation: Calibration

Majority of climate-related models utilize **parameters** whose values might have uncertainties. Some of these parameters cannot be measured directly, and their values must be inferred indirectly based on rather tenuous qualitative arguments or simulation results obtained using other models.

Definition 1.13. *The process of optimizing the model performance by adjusting values of some parameters is called* **model calibration**.

- Model calibration can be carried out using some performance metrics or various qualitative criteria based on **expert knowledge**
- Attention: Calibration (especially based on a limited observation dataset) is a potentially dangerous procedure capable of
  - compromising the physical assumptions the model is based on
  - masking problems in the model or even errors in the model code
  - impairing model's performance for scenarios not used for calibration

Model evaluation: Biases

Definition 1.14. *Systematic (over numerous simulations) deviations of the model results from observations are called* **biases**. *Generally biases can only be reliably detected by comparing observations to the results of* **ensemble simulations** *(multiple simulation runs with perturbed parameters such as forcings or initial conditions).*

- There exist techniques for correction of model biases. As opposed to *a priori* approaches used for model calibration, the bias correction methods belong to *a posteriori* (i.e. postprocessing) procedures
- Since the majority of climate models (especially GCMs) are highly non-linear, avoiding bias corrections by improving the choice of model parameters (better calibration) is rarely possible in practice

Model evaluation: Reproducing the current climate

One of the main tests for general-purpose coupled climate models is an evaluation of the ability to reproduce the current climate using the data for the last 150 years as forcing.

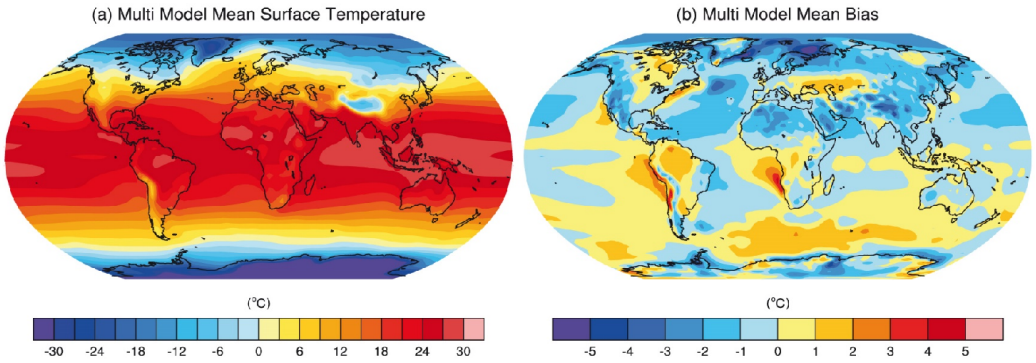


Figure 1.10: Ensemble mean of surface temperature over multiple models (CMIP5, Flato et al. 2013),

Model evaluation: Bias correction

Remark 1.15. *The majority of bias correction techniques work on the assumption that the model captures the main dynamical features of the modeled system (variability, response to forcings, etc.) reasonably well, and that there exists some simple functional relationship between model results and correct values for important unknowns.*

The main steps of bias correction can be summarized as follows:

- Pose a hypothesis on the specific type of the functional relationship F

$$x_{corr}(t) = F(x_{sim}(t), t),$$

  e.g., additive or multiplicative bias, constant or variable in time
- Determine the parameters for the proposed functional relationship using some reference observations
- Correct simulation results using the function with these parameters

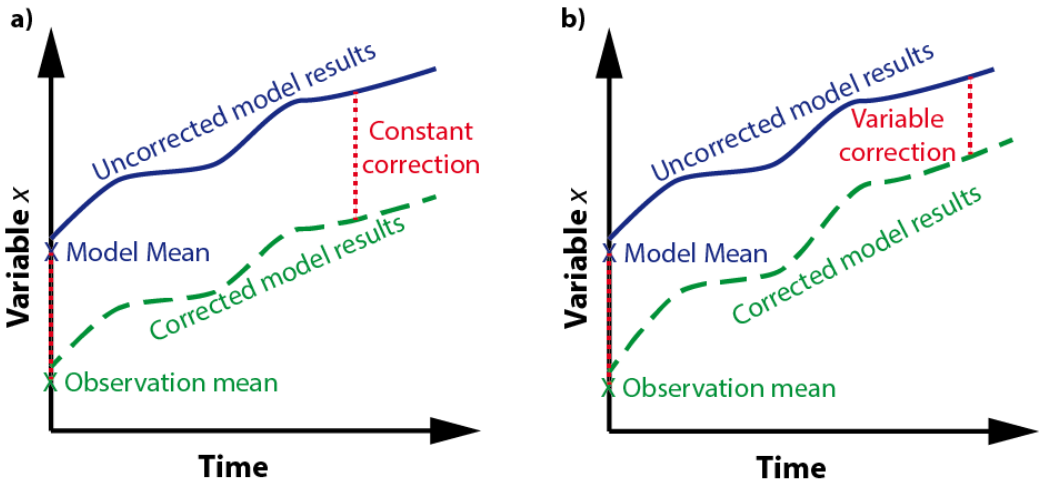


Figure 1.11: Bias correction techniques: a) the mean state has been corrected, b) the mean state and the time derivative have been corrected (Goosse 2015).

Model evaluation: Model intercomparison projects
Model intercomparison projects (MIP) are a rather recent but a very popular development in the climate sciences (not only in climate sciences!) aiming to improve **modeling skill** by simulating a set of pre-specified scenarios using multiple models. The main ideas behind MIPs are:

- Identify and analyze sources of main model biases by comparing similarities and differences between models
- Provide better predictions for important climate development scenarios (biases of multi-model simulations tend to be lower than those of single-model runs)
- Many MIPs in the area of climate modeling (e.g., Coupled Model Intercomparison Project CMIP) have become a standard recurring activity with elaborate sets of input data, formalized evaluation criteria, and dedicated special issues in prominent journals

Questions

- What is the model calibration procedure, and what risks are associated with it?

- What are model biases and how can they be corrected? What assumptions lie behind bias correction techniques?
- What are model intercomparison projects (MIPs), and why are they important?

Data assimilation: Motivation

Definition 1.16. *Data assimilation (DA) is the discipline concerned with incorporating observational data into model simulations* **dynamically, i.e., during simulation**.

- As opposed to the bias correction or calibration techniques, data assimilation cannot be classified as either *a priori* or *a posteriori* procedure
- Data assimilation does not assume a **perfect model** or **error-free observations** but can naturally deal with errors in both
- Historically, data assimilation has been developed as a means to improve numerical weather prediction models (NWP); it is still used heavily in all operational NWP models, and it is still critical for the quality of weather forecasts

Data assimilation: Overview of techniques

The main data assimilation techniques can be classified as follows:

- Nudging
- Optimal interpolation
- Statistical approaches (Kalman filtering)
- Variational methods (3D-Var, 4D-Var)

Data assimilation: Forecast skill

The forecast skill of a model can be quantified based on the **Brier skill score** S defined by

$$S = \frac{e^b - e^f}{e^b},$$

where e^f is the error (in some norm) of the forecast system (e.g. our model), and e^b is the error of some basic forecast method.

Examples of basic forecast methods include:

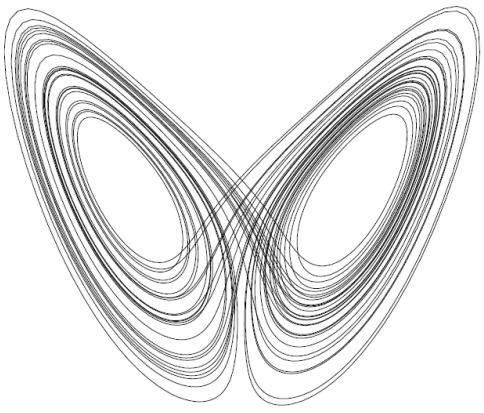
- Climatology (means over long time periods)
- Persistence (no change compared to the initial state)
- Random (e.g. climatology or persistence with added random perturbation with zero average and prescribed variance)

Data assimilation: Lorenz model
A useful tool to investigate sensitivity of weather models to perturbations are **Lorenz equations** – a low-order truncation of spectral equations for atmospheric circulation:

$$x'(t) = \sigma(y - x), \tag{1.12}$$
$$y'(t) = rx - y - xz, \tag{1.13}$$
$$z'(t) = xy - bz, \tag{1.14}$$

where σ, r, and b are some parameters.

Remark 1.17. *These equations are known to generate chaotic deviations from solution trajectories even for small perturbations in initial conditions. This behavior demonstrates the* **predictability limits** *intrinsic to the atmospheric models.*



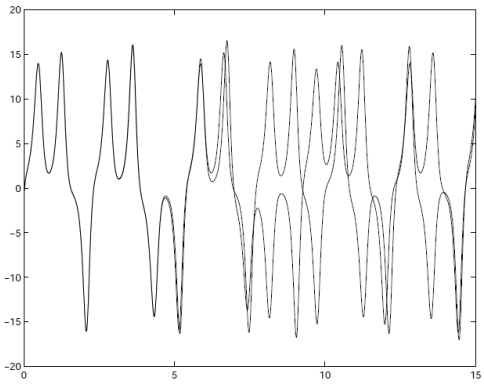


Figure 1.12: Lorenz model with σ= 10, r=28, and b= 8/3: Trajectory in (x, z) space (left) and x(t) for two solutions with slightly different initial conditions (right) (Cushman-Roisin, Beckers 2011).

Data assimilation: Predictability limits
Error accumulation even for arbitrarily small perturbations leads to predictability limits in strongly nonlinear systems. This limit is estimated to be

- One to two weeks for the global atmosphere
- On the order of one month for midlatitude ocean eddies

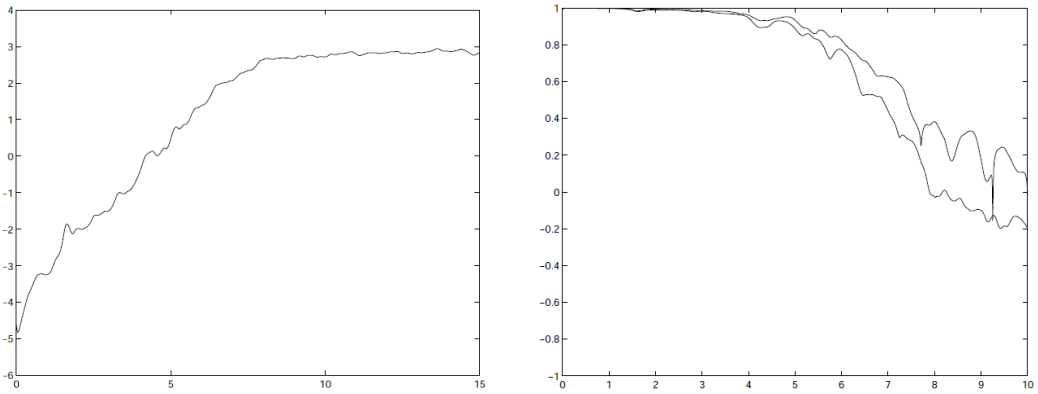


Figure 1.13: Lorenz model: Logarithm of the forecast error (left) and skill score for two forecasts (right) as function of lead time (Cushman-Roisin, Beckers 2011).
A useful measure of predictability is the **autocorrelation** of the solution function for $T \to \infty$ given by

$$\rho(\Delta) = \frac{\frac{1}{T}\int_0^T u(t+\Delta)u(t)dt}{\sqrt{\frac{1}{T}\int_0^T u^2(t)dt}\sqrt{\frac{1}{T}\int_0^T u^2(t+\Delta)dt}}.$$

- This function measures how closely the solution at moment $t + \Delta$ is **on average** related to the solution at moment t. Thus the delay Δ for which $\rho \to 0$ defines the threshold after which the solution is no longer determined by its past values
- For time points farther apart than Δ, the function values are **decorrelated**

- For a purely random signal, the autocorrelation is zero for any $\Delta > 0$

- For the solution of the Lorenz system, a limit of predictability can be expected (cf. Fig. 1.14)

- Note that the system is **deterministic**, i.e. each initial condition determines a unique trajectory. However, this unique trajectory cannot be isolated in practice due to system's sensitivity to even tiniest perturbations

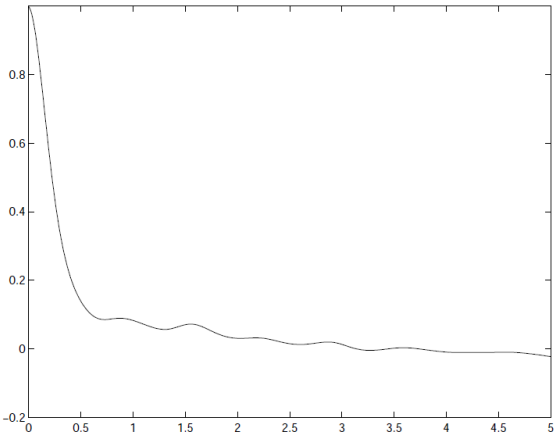


Figure 1.14: Autocorrelation as a function of Δ for solution x(t) of the Lorenz system (Cushman-Roisin, Beckers 2011).

In geophysical circulation systems, solutions are controlled not only by initial conditions but also by boundary conditions and forcings.

Boundary conditions and forcings are often reasonably well-known, thus predictability limits depend mainly on the relative importance of boundary vs. initial conditions.

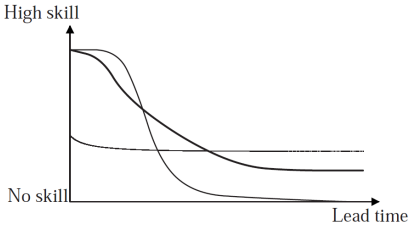


Figure 1.15: Predictability for systems dominated by boundary condition or forcings (flat line), initial condition (steep line), and mixed situations (intermediate line) (Cushman-Roisin, Beckers 2011).

For tidal dynamics in shallow seas, long simulation times with little loss of accuracy are possible since the skill is mainly constrained by the forcings.

For the global atmosphere, the initial conditions are usually constructed with care, and the initial skill is high but rapidly drops with time.

Data assimilation: Nudging

Idea: If 'better' values of simulated fields are known, push the simulation towards these values.

Denote by $\boldsymbol{\mathcal{Q}}(\boldsymbol{x}, t)$ the model equations and consider

$$\boldsymbol{x}'(t) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}, t).$$

If the observation $\boldsymbol{y}$ can be represented on the same grid as the solution $\boldsymbol{x}$, we can add a linear correction term to our system by

$$\boldsymbol{x}'(t) = \boldsymbol{\mathcal{Q}}(\boldsymbol{x}, t) + \boldsymbol{K}(\boldsymbol{y} - \boldsymbol{x}).$$

For nudging, $\boldsymbol{K}$ is a diagonal matrix given by $\boldsymbol{K}$ =diag$(1/\tau)$, where τ is the **relaxation** time controlling the 'strength' of nudging. For values of τ large compared to the time scale of the system, the nudging is gentle. One popular choice of τ is given by

$$1/\tau = K \exp\left(-(t - t_0)^2/T^2\right),$$

where t_0 is the time point with available observations and T the time scale allocated to this observation to significantly affect the simulation process.

Data assimilation: Sequential assimilation

The nudging approach is useful but ad-hoc. Next, we consider sequential assimilation techniques based on statistical methods. They **re-initialize** the model by incorporating observations into the model results.

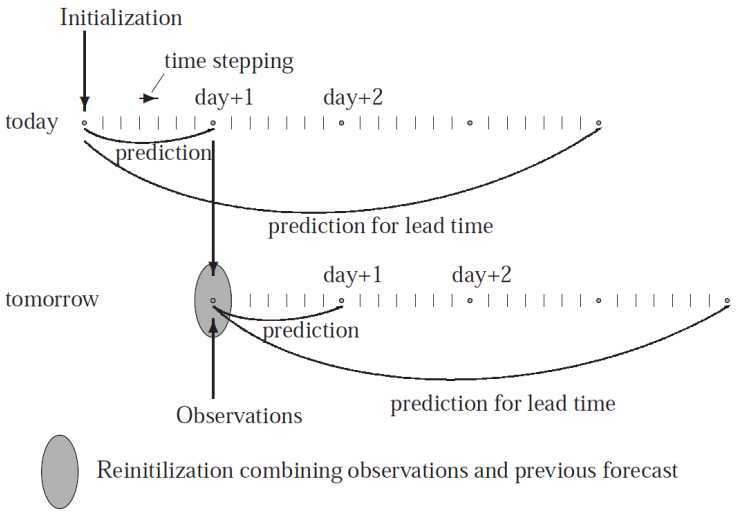


Figure 1.16: Illustration of reinitialization: A new initial condition for the model (called **analysis**) is created by combining the forecast and observations of the real system (Cushman-Roisin, Beckers 2011).

Data assimilation: Optimal interpolation for two values

Example 1.18. *Consider two approximations T_1 and T_2 to the 'true' value T^t of some unknown (e.g., temperature) with corresponding errors ϵ_1 and ϵ_2, whose error variances are known:*

$$T_1 = T^t + \epsilon_1, \quad T_2 = T^t + \epsilon_2.$$

We assume our approximations to be **unbiased**, *i.e.* $\langle \epsilon_1 \rangle = \langle \epsilon_2 \rangle = 0$, *where $\langle \cdot \rangle$ denotes the mean value.*
We seek weights w_1, w_2 to construct a 'better' approximation using T_1, T_2

$$T = w_1 T_1 + w_2 T_2 = (w_1 + w_2)T^t + w_1\epsilon_1 + w_2\epsilon_2.$$

We obviously have

$$\langle T \rangle = (w_1 + w_2)T^t,$$

thus for the estimate of T^t to be unbiased, $w_1 + w_2 = 1$ must hold.
Then the unbiased estimate (analysis) T^a is given by

$$T^a = (1 - w_2)T_1 + w_2 T_2 = T_1 + w_2(T_2 - T_1)$$

and its error

$$\epsilon = T_a - T_t = (1 - w_2)\epsilon_1 + w_2\epsilon_2.$$

The error has zero mean

$$\langle \epsilon \rangle = (1 - w_2)\langle \epsilon_1 \rangle + w_2 \langle \epsilon_2 \rangle = 0,$$

but its variance is generally non-zero

$$\langle \epsilon^2 \rangle = \langle (T_a - T_t)^2 \rangle = (1 - w_2)^2 \langle \epsilon_1^2 \rangle + w_2^2 \langle \epsilon_2^2 \rangle + 2(1 - w_2) w_2 \langle \epsilon_1 \epsilon_2 \rangle.$$

If measurements T_1, T_2 are uncorrelated, then $\langle \epsilon_1 \epsilon_2 \rangle = 0$ holds giving

$$\langle \epsilon^2 \rangle = (1 - w_2)^2 \langle \epsilon_1^2 \rangle + w_2^2 \langle \epsilon_2^2 \rangle = (\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle) w_2^2 - 2\langle \epsilon_1^2 \rangle w_2 + \langle \epsilon_1^2 \rangle,$$

which is a quadratic function of w_2 with a positive leading coefficient.
The solution minimizing the error variance $\langle \epsilon^2 \rangle$ is then

$$w_2 = \frac{\langle \epsilon_1^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \quad \text{with} \quad T^a = \frac{\langle \epsilon_2^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} T_1 + \frac{\langle \epsilon_1^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} T_2.$$

Note that the analysis error is at least as good as that of either T_1 or T_2

$$\langle \epsilon^2 \rangle = \frac{\langle \epsilon_1^2 \rangle \langle \epsilon_2^2 \rangle}{\langle \epsilon_1^2 \rangle + \langle \epsilon_2^2 \rangle} \leq \min\{\langle \epsilon_1^2 \rangle, \langle \epsilon_2^2 \rangle\}.$$

The same result can be alternatively obtained as the solution to the minimization problem

$$J(T) \to \min \quad \text{with} \quad J(T) = \frac{(T - T_1)^2}{2\langle \epsilon_1^2 \rangle} + \frac{(T - T_2)^2}{2\langle \epsilon_2^2 \rangle}.$$

Questions

- What is the idea behind data assimilation? Where does it fit between the calibration and bias correction?
- What is the forecast skill, and how can it be quantified?
- What is a predictability limit, and what values is it estimated to have for weather prediction models?
- How can one measure the predictability limit of a given model?
- Discuss the differences in the long time behavior for systems that are
  - controlled by the initial conditions
  - controlled by the boundary data and forcings
- Give the general idea of data assimilation by means of nudging
- What is the main idea behind the sequential assimilation?
- Given two approximation to the true solution (e.g. simulation and observation), how does one construct a 'better' approximation?

Data assimilation: Quick guide to statistics

- The covariance between two variables x_i and x_j is defined as

$$\mathrm{cov}(x_i, x_j) = \langle (x_i - \langle x_i \rangle)\,(x_j - \langle x_j \rangle) \rangle.$$

  Given a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^T$, the covariances can be arranged into a covariance matrix, $\mathrm{cov}(\boldsymbol{x})$, such that $\mathrm{cov}(\boldsymbol{x})_{ij} = \mathrm{cov}(x_i, x_j)$. Equivalently

$$\mathrm{cov}(\boldsymbol{x}) = \langle (\boldsymbol{x} - \langle \boldsymbol{x} \rangle)\,(\boldsymbol{x} - \langle \boldsymbol{x} \rangle)^T \rangle.$$

  Covariance matrices are symmetric positive semi-definite.

- Given two vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, one can calculate in a similar way the cross-covariance matrix

$$\mathrm{cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle (\boldsymbol{x}_i - \langle \boldsymbol{x}_i \rangle)\,(\boldsymbol{x}_j - \langle \boldsymbol{x}_j \rangle)^T \rangle.$$

  Cross-covariance matrices of uncorrelated vectors are filled with zeros.

Data assimilation: Optimal interpolation

Similarly to the nudging problem, we denote by $\boldsymbol{x}^f$ the model solution (forecast) vector and by $\boldsymbol{y}$ the observations. However, the vector dimensions can now be different $\boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{y} \in \mathbb{R}^p$.

Then the corrected solution (analysis) vector $\boldsymbol{x}^f$ can be expressed as

$$\boldsymbol{x}^a = \boldsymbol{x}^f + \boldsymbol{K}(\boldsymbol{y} - \boldsymbol{\mathcal{H}}\boldsymbol{x}^f),$$

where $\mathcal{H}$ is called the **observation operator** and $\boldsymbol{K}$ the **gain matrix**.

- The role of the observation operator $\boldsymbol{\mathcal{H}} : \mathbb{R}^m \mapsto \mathbb{R}^p$ is to map the solution vector $\boldsymbol{x}$ to the 'observation space' of vector $\boldsymbol{y}$ to make both vectors *comparable*. If vector $\boldsymbol{y}$ denotes observations at some locations, $\boldsymbol{\mathcal{H}}$ can be, e.g., an operator interpolating solution vector $\boldsymbol{x}$ to these locations

- Gain matrix $\boldsymbol{K} \in \mathbb{R}^{m \times p}$ assigns 'weights' to single observations, and the choice of this matrix is critical for the 'quality' of the analysis

From now on, we assume the observation operator $\boldsymbol{\mathcal{H}}$ is linear and denote it by $\boldsymbol{H} \in \mathbb{R}^{p\times m}$.
We write $\boldsymbol{\epsilon}^f = \boldsymbol{x}^f - \boldsymbol{x}^t, \boldsymbol{\epsilon}^o = \boldsymbol{y} - \boldsymbol{y}^t$ for the forecast and observation error, respectively, and get $\langle \boldsymbol{\epsilon}^f \rangle = \langle \boldsymbol{\epsilon}^o \rangle = 0$ by assuming $\boldsymbol{x}^f, \boldsymbol{y}$ to be unbiased.
We also assume that

$$\boldsymbol{K}(\boldsymbol{y}^t - \boldsymbol{H}\boldsymbol{x}^t) = 0, \tag{1.15}$$

i.e. the projection of the 'true' solution exactly fits 'true' observations.
The observation error covariance matrix

$$\boldsymbol{R} = \langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^{oT} \rangle$$

contains the error variance of each observation on the diagonal whereas the off-diagonal terms represent the correlations of corresponding two observations. Off-diagonal terms may become non-zero when $\boldsymbol{y}$ represents, e.g., satellite observations exhibiting correlated errors.
The analysis step can then be expressed as

$$\boldsymbol{x}^a = \boldsymbol{x}^f + \boldsymbol{K}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^f)$$

or $$\underbrace{\boldsymbol{x}^t + \boldsymbol{\epsilon}^a}_{\boldsymbol{x}^a =} = \underbrace{\boldsymbol{x}^t + \boldsymbol{\epsilon}^f}_{\boldsymbol{x}^f =} + \boldsymbol{K}(\boldsymbol{\epsilon}^o - \boldsymbol{H}\boldsymbol{\epsilon}^f) + \underbrace{\boldsymbol{K}(\boldsymbol{y}^t - \boldsymbol{H}\boldsymbol{x}^t)}_{=0 \text{ (c.f. (1.15))}}$$

Hence the error of the analysis reads

$$\boldsymbol{\epsilon}^a = \boldsymbol{\epsilon}^f + \boldsymbol{K}(\boldsymbol{\epsilon}^o - \boldsymbol{H}\boldsymbol{\epsilon}^f). \tag{1.16}$$

The goal is now to minimize the analysis error norm

$$\|\boldsymbol{\epsilon}^a\| = \sqrt{\langle \boldsymbol{\epsilon}^{a^T} \boldsymbol{\epsilon}^a \rangle}$$

with respect to gain matrix $\boldsymbol{K}$ controlling the quality of the analysis.
The solution to this problem is called the **optimal interpolation**.
We start by constructing the error covariance matrix $\langle \boldsymbol{\epsilon}^a \boldsymbol{\epsilon}^{aT} \rangle$ of the analysis by multiplying (1.16) with its transposed and taking the average:

$$\begin{aligned}\langle \boldsymbol{\epsilon}^a \boldsymbol{\epsilon}^{aT} \rangle &= \langle \boldsymbol{\epsilon}^f \boldsymbol{\epsilon}^{f^T} \rangle + \boldsymbol{K}\langle (\boldsymbol{\epsilon}^o - \boldsymbol{H}\boldsymbol{\epsilon}^f)\boldsymbol{\epsilon}^{f^T} \rangle + \langle \boldsymbol{\epsilon}^f (\boldsymbol{\epsilon}^{o^T} - \boldsymbol{\epsilon}^{f^T}\boldsymbol{H}^T)\rangle \boldsymbol{K}^T \\ &+ \boldsymbol{K}\langle (\boldsymbol{\epsilon}^o - \boldsymbol{H}\boldsymbol{\epsilon}^f)(\boldsymbol{\epsilon}^{oT} - \boldsymbol{\epsilon}^{f^T}\boldsymbol{H}^T)\rangle \boldsymbol{K}^T.\end{aligned}$$

Denoting error covariance matrices for the forecasted and analyzed solutions by

$$\boldsymbol{P}^f = \langle \boldsymbol{\epsilon}^f \boldsymbol{\epsilon}^{f^T} \rangle, \quad \boldsymbol{P}^a = \langle \boldsymbol{\epsilon}^a \boldsymbol{\epsilon}^{a^T} \rangle,$$

we can expand the expressions and assume that observational and model errors are uncorrelated, i.e. $\langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^{f^T} \rangle = 0$ (this assumption is justified by different origins of information), to obtain

$$\boldsymbol{P}^a = \boldsymbol{P}^f - \boldsymbol{K}\boldsymbol{H}\boldsymbol{P}^f - \boldsymbol{P}^f\boldsymbol{H}^T\boldsymbol{K}^T + \boldsymbol{K}\left(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T\right)\boldsymbol{K}^T.$$

On the other hand, note that

$$\|\boldsymbol{\epsilon}^a\|^2 = \langle \boldsymbol{\epsilon}^{a^T} \boldsymbol{\epsilon}^a \rangle = \operatorname{trace}\left(\boldsymbol{P}^a\right)$$

is a quadratic form in terms of entries of matrix $\boldsymbol{K}$. In addition, matrix $\boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T$ multiplying the quadratic terms of this form is symmetric semi-positive definite, thus the minimization problem has a unique solution

$$\boldsymbol{K}_* = \boldsymbol{P}^f\boldsymbol{H}^T\left(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T\right)^{-1}$$

called the **Kalman gain matrix** with analysis and covariance given by

$$\boldsymbol{x}^a = \boldsymbol{x}^f + \boldsymbol{K}_*(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^f), \qquad \boldsymbol{P}^a = \left(\boldsymbol{I} - \boldsymbol{K}_*\boldsymbol{H}\right)\boldsymbol{P}^f.$$

Data assimilation: 3D-Var method

Remark 1.19. *Computing $\boldsymbol{K}_*$ involves explicitly inverting a potentially very large matrix $\boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}^f\boldsymbol{H}^T \in \mathbb{R}^{p\times p}$. Alternatively, the same $\boldsymbol{x}^a$ can be obtained by minimizing quadratic* **cost functional** *$J(\boldsymbol{x})$ given by*

$$J(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{x}^f\right)^T\left(\boldsymbol{P}^f\right)^{-1}\left(\boldsymbol{x} - \boldsymbol{x}^f\right) + \frac{1}{2}\left(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}\right)^T\boldsymbol{R}^{-1}\left(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}\right).$$

- This approach is called **3D-Var** (due to the variational minimization problem)
- The advantage of this method is the ability to compute an approximation to $\boldsymbol{x}^a$ *iteratively*

Data assimilation: Kalman filtering

In the optimal interpolation problem defined above, the model providing forecast only enters implicitly via the forecast error covariance matrix $\boldsymbol{P}^f$.

Idea: One can utilize more information from the model by using it to propagate $\boldsymbol{P}^f$ as well – this approach is called **Kalman filtering**.

We assume that between assimilation cycles n and $n+1$, the model advances the solution vector in time according to

$$\boldsymbol{x}_{n+1} = \boldsymbol{\mathcal{M}}(\boldsymbol{x}_n) + \boldsymbol{f}_n \ + \boldsymbol{\eta}_n.$$

Here $\boldsymbol{\eta}_n$ takes into account errors introduced by the model, and $\boldsymbol{f}_n$ includes the external forcings. Operator $\boldsymbol{\mathcal{M}}(\boldsymbol{x}_n)$ represents the total model effect on $\boldsymbol{x}_n$ between assimilation cycles.
Assuming $\boldsymbol{\mathcal{M}}$ to be linear and denoting it by matrix $\boldsymbol{M}$ we note that the true state evolves without modeling errors according to

$$\boldsymbol{x}^t_{n+1} = \boldsymbol{M}\boldsymbol{x}^t_n + \boldsymbol{f}_n,$$

so that the forecast error $\boldsymbol{\epsilon}^f = \boldsymbol{x}^f - \boldsymbol{x}^t$ satisfies

$$\boldsymbol{\epsilon}^f_{n+1} = \boldsymbol{M}\boldsymbol{\epsilon}^a_n + \boldsymbol{\eta}_n. \tag{1.17}$$

Multiplying (1.17) by its transposed, averaging, and assuming errors of different origins to be uncorrelated, we can advance the forecast error covariance matrix $\boldsymbol{P}^f$ in time using the so-called Lyapunov equation

$$\boldsymbol{P}^f_{n+1} = \boldsymbol{M}\boldsymbol{P}^a_n\boldsymbol{M}^T \ + \ \langle \boldsymbol{\eta}_n \, \boldsymbol{\eta}^T_n \rangle,$$

where $\langle \boldsymbol{\eta}_n \, \boldsymbol{\eta}^T_n \rangle$ represents the model error covariance matrix.

Remarks 1.20.

- *The initial value of* $\boldsymbol{P}^a$ *can be computed from the initial condition*

$$\boldsymbol{P}_0 = \langle (\boldsymbol{x}_0 - \boldsymbol{x}_0^t)(\boldsymbol{x}_0 - \boldsymbol{x}_0^t)^T \rangle$$

- *Matrix* $\boldsymbol{M}$ *does not have to be computed explicitly: Since*

$$\boldsymbol{M}\boldsymbol{P}_n^a\boldsymbol{M}^T = \boldsymbol{M}(\boldsymbol{M}\boldsymbol{P}_n^a)^T,$$

  $\boldsymbol{P}^a$ *can be evolved by applying the model operator to its columns*

- *Kalman filtering contains other data assimilation methods as special cases:*
  - *prescribing the forecast error covariance matrix instead of propagating it results in the optimal interpolation method*
  - *if the model and observation error matrices are given and diagonal, one gets a nudging scheme*

Initialization:
$\boldsymbol{x}_0^a = \boldsymbol{x}_0$
$\boldsymbol{P}_0^a = \boldsymbol{P}_0$

↓

Loop over $n = 1, 2, \ldots$

Forecast:
$\boldsymbol{x}_{n+1}^f = \boldsymbol{\mathcal{M}}(\boldsymbol{x}_n^a) + \boldsymbol{f}_n$
$\boldsymbol{P}_{n+1}^f = \boldsymbol{M}_n\boldsymbol{P}_n^a\boldsymbol{M}_n^T + \langle \boldsymbol{\eta}_n\,\boldsymbol{\eta}_n^T \rangle$

↓

Analysis:
$$\boldsymbol{K}_{n+1} = \boldsymbol{P}_{n+1}^f\boldsymbol{H}_{n+1}^T\left(\boldsymbol{H}_{n+1}\boldsymbol{P}_{n+1}^f\boldsymbol{H}_{n+1}^T + \boldsymbol{R}_{n+1}\right)^{-1}$$
$$\boldsymbol{x}_{n+1}^a = \boldsymbol{x}_{n+1}^f + \boldsymbol{K}_{n+1}(\boldsymbol{y}_{n+1} - \boldsymbol{H}_{n+1}\boldsymbol{x}_{n+1}^f)$$
$$\boldsymbol{P}_{n+1}^a = \boldsymbol{P}_{n+1}^f - \boldsymbol{K}_{n+1}\boldsymbol{H}_{n+1}\boldsymbol{P}_{n+1}^f$$

Data assimilation: 4D-Var methods

Kalman filtering (and optimal interpolation in general) assimilate observations at a given time instant. After analysis (and reinitialization),

the solution may jump to a new state and trigger model instabilities.

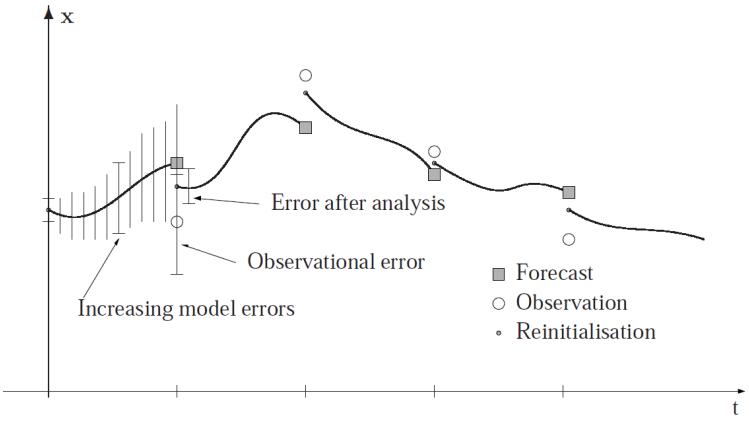


Figure 1.17: Kalman filtering reduces the error but leads to a model trajectory interrupted at each assimilation cycle (Cushman-Roisin, Beckers 2011).

The main reason for this problem is the approximation error when propagating the error covariance matrix due to $\boldsymbol{M} \approx \boldsymbol{\mathcal{M}}$.

Idea: Assimilate observations distributed over multiple time points and choose the solution trajectory optimal with respect to these data.

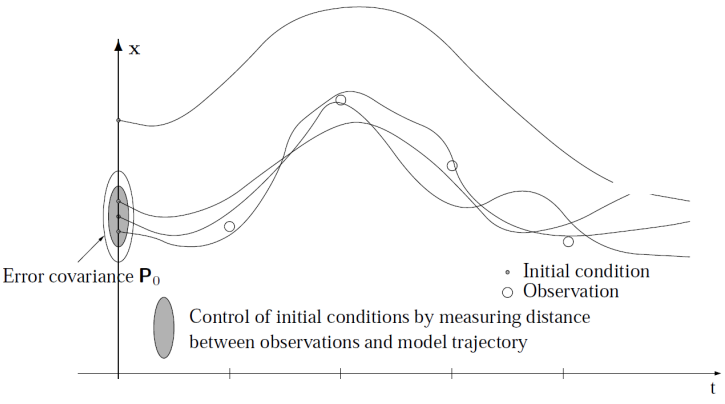


Figure 1.18: 4D-Var methods use adjoint approach allowing to compute the model trajectory that optimally fits observations over a given time interval (Cushman-Roisin, Beckers 2011).

- This methodology is based on solving an **inverse** or **adjoint** problem and is called **4D-Var** data assimilation
- It is powerful but computationally very expensive for real-life problems

To achieve this, we consider observational datasets $\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ at corresponding time instants and denote by $\boldsymbol{x}_0^b$ some initial background state (e.g. provided by a large-scale model) that we want to stay close to. Solution

trajectory specified by $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ is then obtained by minimizing cost functional $J(\boldsymbol{x})$ given by

$$J(\boldsymbol{x}) = \frac{1}{2}\sum_{n=0}^{N}\left(\boldsymbol{H}_n\boldsymbol{x}_n - \boldsymbol{y}_n\right)^T \boldsymbol{R}_n^{-1}\left(\boldsymbol{H}_n\boldsymbol{x}_n - \boldsymbol{y}_n\right) + \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^T \boldsymbol{P}_0^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right) \tag{1.18}$$

subject to **constraints**

$$\boldsymbol{x}_{n+1} = \boldsymbol{\mathcal{M}}(\boldsymbol{x}_n) + \boldsymbol{f}_n, \quad n = 0, 1, \ldots, N-1$$

or, once again resorting to a linearized formulation,

$$\boldsymbol{x}_{n+1} = \boldsymbol{M}\boldsymbol{x}_n + \boldsymbol{f}_n, \quad n = 0, 1, \ldots, N-1. \tag{1.19}$$

A standard way to solve constrained minimization problem (1.18), (1.19) is the **Lagrange multiplier** method that formulates an **unconstrained** minimization problem for the extended cost functional given by

$$\begin{aligned} J(\boldsymbol{x}) = \frac{1}{2}\sum_{n=0}^{N}&\left(\boldsymbol{H}_n\boldsymbol{x}_n - \boldsymbol{y}_n\right)^T \boldsymbol{R}_n^{-1}\left(\boldsymbol{H}_n\boldsymbol{x}_n - \boldsymbol{y}_n\right) + \sum_{n=0}^{N-1}\boldsymbol{\lambda}_n^T(\boldsymbol{x}_{n+1} - \boldsymbol{M}\boldsymbol{x}_n - \boldsymbol{f}_n) \\ &+ \frac{1}{2}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right)^T \boldsymbol{P}_0^{-1}\left(\boldsymbol{x}_0 - \boldsymbol{x}_0^b\right). \end{aligned} \tag{1.20}$$

The minimization is then carried out with respect to solution vectors $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ and **Lagrange multipliers** $\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_{N-1}$.

Questions

- If the cross-correlation matrix of two vectors is zero, what does this signify?
- Given the optimal interpolation problem

  $$\boldsymbol{x}^a = \boldsymbol{x}^f + \boldsymbol{K}(\boldsymbol{y} - \boldsymbol{\mathcal{H}}\boldsymbol{x}^f),$$

  what is the role of the observation operator $\boldsymbol{\mathcal{H}}$?

  ... of the gain matrix $\boldsymbol{K}$?

- How is the 3D-Var method related to the optimal interpolation problem? What are the advantages of the variational approach?
- What is the idea behind the Kalman filtering?
- What is the main motivation for the 4D-Var data assimilation?
- Why are 4D-Var methods not as frequently used as their advantages would suggest?

2 Compartments of the climate system and their modeling

2.1 Energy balance models

Earth's energy budget
At the top of the atmosphere, the total solar irradiance (TSI) is $S_o = 1368$ W/m^2. This amount is not constant but can vary due to a number of causes that can be summarized as follows:

Phenomenon	Effect	Periodicity
Changes in Earth's orbit eccentricity	$\approx 0.1\%$	100.000 – 400.000 years
Changes in the angle of ecliptic	$\approx 0.5\%$	41.000 years
Precession of Earth's axis	$\approx 4\%$	19.000 – 23.000 years
Sunspots	$\approx 0.1\%$	9-10 years

Table 2.1: Astronomical factors affecting Earth's energy budget.

The fraction of the incoming solar radiation that is reflected is called the **albedo** of the Earth or planetary albedo (α_p), thus the amount of solar energy per unit of area absorbed by the Earth ($A^{\downarrow}$) is given by

$$A^{\downarrow} = (1 - \alpha_p)S_o.$$

For present-day conditions, α_p has a value of about 0.3.
At the top of Earth's atmosphere, the average total incoming solar energy per unit of time is the TSI times the surface that intercepts the solar rays (πR^2 with $R = 6371$ km).

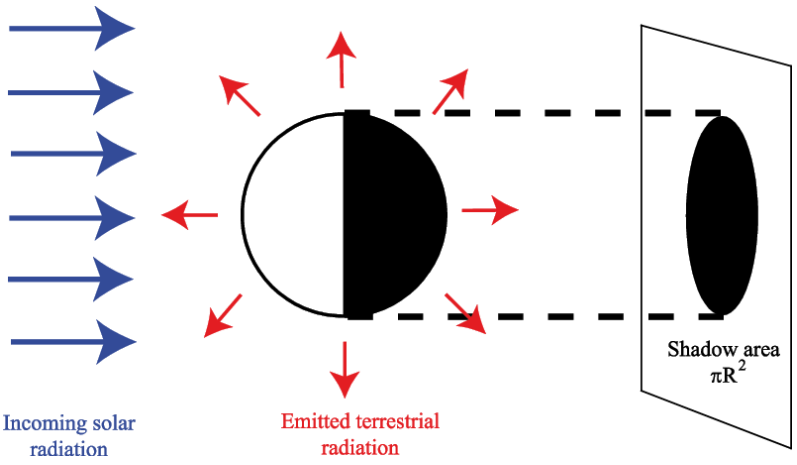


Figure 2.1: Schematic of the incoming/outgoing radiation (Hartmann 1994). At the top of the atmosphere, the solar energy (shortwave radiation) is balanced by the energy emitted by the Earth (longwave radiation).

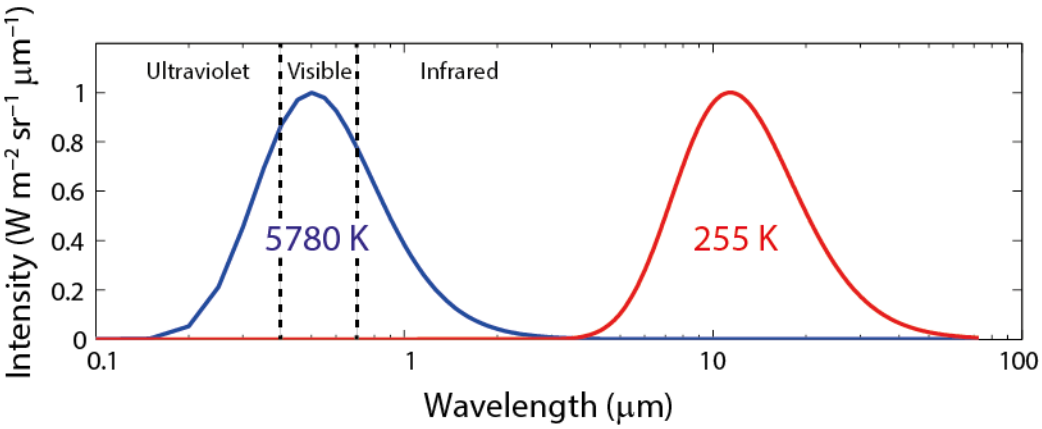


Figure 2.2: Normalized **blackbody** spectra for representative temperatures of the Sun (blue, 5780 K) and the Earth (red, 255 K) (Goosse 2015).

- The total amount of energy that is emitted by a 1 m^2 surface per unit of time by the Earth at the top of the atmosphere ($A^{\uparrow}$) can be computed following Stephan-Boltzmann's law

  $$A^{\uparrow} = \sigma T_e^4,$$

  where T_e is the effective emission temperature of the Earth, and $\sigma = 5.67 \times 10^{-8} W m^{-2} K^{-4}$ is the Stefan-Boltzmann constant

- Given the total Earth surface of $4\pi R^2$, the balance becomes

$$\pi R^2(1-\alpha_p)S_o = 4\pi R^2 \sigma T_e^4 \quad \Leftrightarrow \quad \frac{1}{4}(1-\alpha_p)S_o = \sigma T_e^4 \tag{2.1}$$

- Solving equation (2.1) for T_e gives

$$T_e = \sqrt[4]{\frac{1}{4\sigma}(1-\alpha_p)S_o} = 255\text{K} = \text{-}18^\circ\text{C}$$

Greenhouse effect

Earth's atmosphere is nearly transparent to visible light but opaque across most of the infrared part of the electromagnetic spectrum because of minor air constituents (water vapor, carbon dioxide, methane, and ozone). This phenomenon is called the **greenhouse effect**.

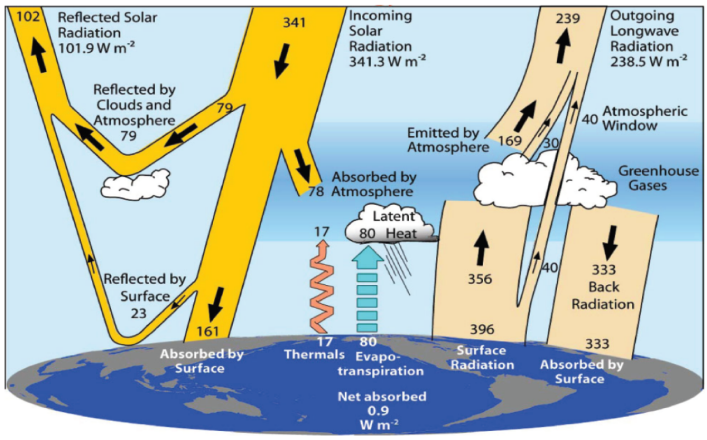


Figure 2.3: Schematic of Earth's energy budget (Trenberth et al. 2009).

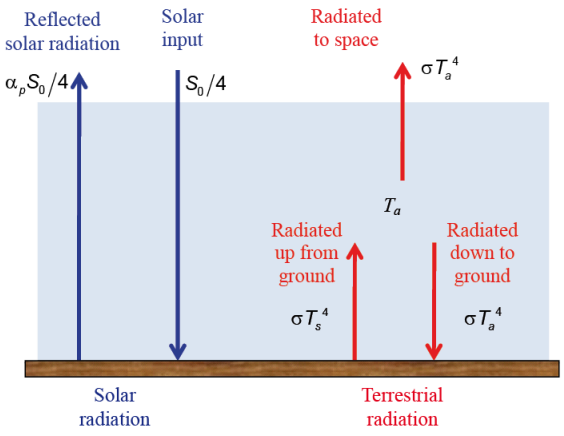


Figure 2.4: Heat balance of the Earth with the atmosphere represented by a single layer transparent to solar radiation and opaque to infrared radiation (Goosse 2015).

Representing the atmosphere by a single homogeneous layer of temperature T_a totally transparent to the solar radiation and totally opaque to the infrared radiation, the heat balance at the top of the atmosphere can be written as

$$\frac{1}{4}(1-\alpha_p)S_o = \sigma T_a^4 = \sigma T_e^4,$$

whereas the heat balance at the Earth surface has the form

$$\sigma T_s^4 = \frac{1}{4}(1-\alpha_p)S_o + \sigma T_a^4,$$

corresponding to $T_s = \sqrt[4]{2}T_e \approx 1.19T_e = 303$ K $= 30°$ C.

- More precise estimates of the radiative balance of the Earth must account for absorptions and reemissions by various atmospheric layers, selective absorptions in specific frequency bands characteristic for each constituent, and the contribution of non-radiative exchanges

- Since the interplay of different factors affecting Earth's energy budget is very complex, one can try to determine the **effective** parameters describing the current energy state (c.f. model calibration). We modify equation (2.1) by introducing **emissivity**[2] ε obtaining

$$\frac{1}{4}(1-\alpha_p)S_o = \varepsilon\sigma T^4$$

and calculate the value of ε based on the current mean Earth surface temperature $T_m = 287.7\text{K} = 14.7°\text{C}$ resulting in

$$\varepsilon \approx 0.62$$

[2]Emissivity $0 \leq \varepsilon \leq 1$ of a body is the ratio between the energy emitted and the energy emitted by the blackbody at the same temperature

A zero-dimensional Energy Balance Model

- Assuming that the Earth's surface has a constant heat capacity C (the amount of energy per unit of area needed to raise the temperature by one degree), one can formulate an instationary version of (2.1) as

$$CT'(t) = \frac{1}{4}(1 - \alpha_p)S_o - \sigma T^4$$

- The assumption that the planetary albedo does not depend on the temperature is not realistic since cold temperatures usually result in an increase in the snow/ice cover and thus in albedo. Introducing temperature-dependent albedo $\alpha_p(T)$ that has the following property:

$$\alpha_p(T) \approx \begin{cases} 0.7 & \text{if } T < 250\text{K}, \\ 0.3 & \text{if } T \geq 280\text{K}, \end{cases}$$

one can take an expression that provides a smooth transition, e.g.,

$$\alpha_p(T) = 0.5 - 0.2 \tanh\left(\frac{T - 265}{10}\right).$$

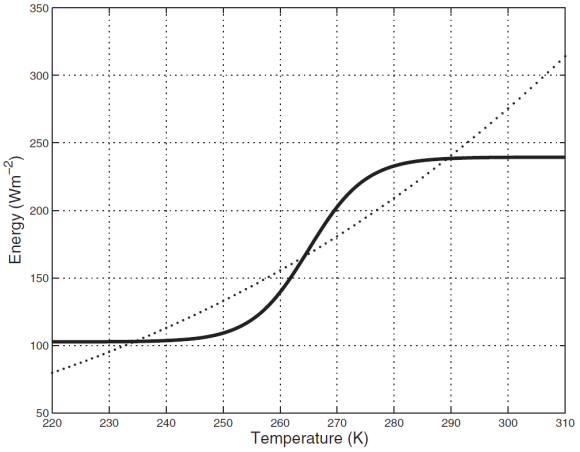


Figure 2.5: Equilibria of the zero-dimensional EBM (Kaper, Engler 2013).

- Fig. 2.5 plots the incoming solar radiation (solid line) taken as $\frac{1}{4}(1 - \alpha_p)S_o$ and the emitted infrared radiation (dotted line) given by $\varepsilon\sigma T^4$ for $\varepsilon = 0.6$
- Intersection points are equilibria of the model where $T'(t) = 0$
- Note that the equilibrium at $T_2^* \approx 263$K is **unstable**, i.e., any perturbation drives $T(t)$ away from T_2^*
- The rightmost equilibrium at $T_3^* \approx 288$K corresponds to the current climate, whereas $T_1^* \approx 235$K represents the '**snowball Earth**'

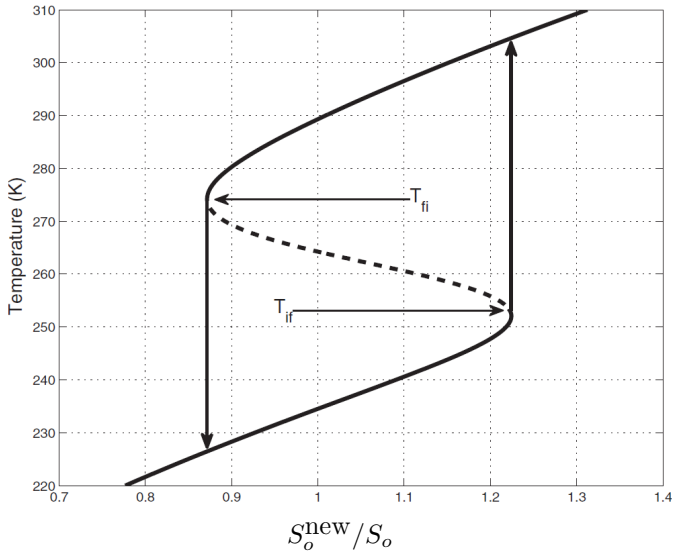


Figure 2.6: Bifurcation points of the zero-dimensional EBM (Kaper, Engler 2013).

- Increasing S_o results in T_2^* and T_3^* moving closer to each other and, eventually merging and disappearing
- Similar property holds for T_1^* and T_2^* when S_o decreases
- This phenomenon signifies qualitative changes in the character of the system and is called **bifurcation**

A one-dimensional Energy Balance Model

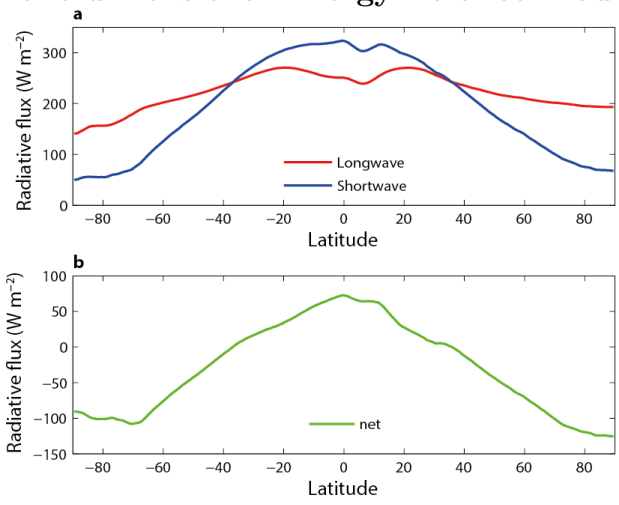


Figure 2.7: Top: annual zonal mean of the absorbed solar radiation and the outgoing longwave radiation at the top of the atmosphere [W/m^2]; bottom: the difference (net mean) (Goosse 2015).

- Until now, radiative energy fluxes were averaged over the total Earth surface
- However, the absorbed and the emitted radiation strongly depend on the geographic latitude φ (Fig. 2.7)

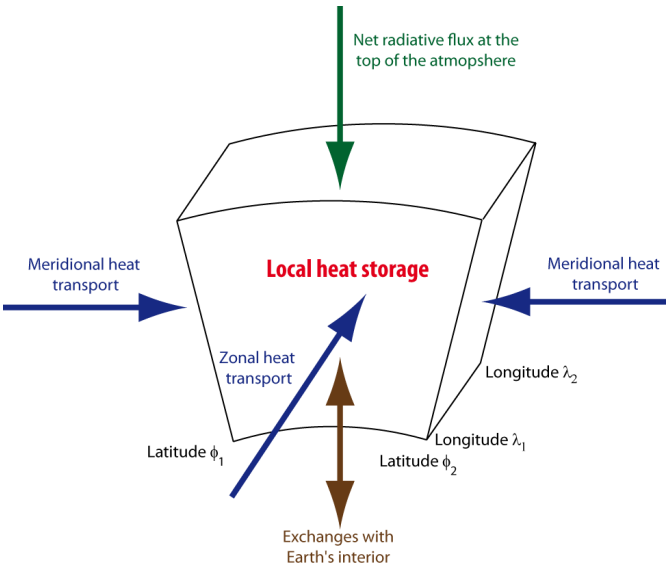


Figure 2.8: Schematic of local energy balance for an elementary volume of Earth's surface (Goosse 2015).

- This radiative flux imbalance is almost entirely compensated for by the **energy transport**
- The main energy transport media are the atmosphere and the ocean with the former playing a much larger role

The total energy consists of **sensible heat** cT, potential energy gz, **latent heat**[3] Lq, and kinetic energy $0.5|\boldsymbol{u}|^2$; it can be expressed as

$$E = cT + gz + Lq + 0.5|\boldsymbol{u}|^2,$$

where c is the specific heat capacity of the medium (air or water), g the gravity acceleration, z the height above a reference level, L the specific latent heat, q the specific humidity (mass of water vapor per unit of dry air), and $\boldsymbol{u}$ the velocity vector (mostly without the vertical component).

Name	Formula	Amount $[10^6 \times Jm^{-2}]$	Percentage of total
Sensible heat	cT	1800	70.2
Potential energy	gz	700	27.3
Latent heat	Lq	64	2.5
Kinetic energy	$0.5\|\boldsymbol{u}\|^2$	1.2	0.05

- The meridional flux of sensible heat in the atmosphere F can be expressed (we assume a specific parametrization here!)

$$F = -\rho c K(\varphi) \frac{1}{R} \frac{\partial T}{\partial \varphi},$$

[3]Note that this term is formulated in the form used in atmospheric modeling although the latent heat of sea ice melt can be also a part of the energy budget in the ocean

where ρ is the air density and $K(\varphi)$ the effective zonal diffusivity dependent on latitude φ

- Using this expression, the one-dimensional energy balance model for a layer of thickness h can be formulated[4] as follows:

$$h\rho c\frac{\partial T}{\partial t} = \frac{h}{R\cos\varphi}\frac{\partial}{\partial\varphi}\left(\rho cK(\varphi)\frac{1}{R}\cos\varphi\frac{\partial T}{\partial\varphi}\right) + \frac{1-\alpha(\varphi)}{4}S(\varphi) - \varepsilon(\varphi)\sigma T^4,$$

where the effective diffusivity K, the albedo α, and the emissivity ε may be functions of latitude. The (mainly shortwave) incoming radiation $S(\varphi)$ also depends on the latitude.

Questions

- What main types of radiation contribute to Earth's energy budget?
- What is a blackbody and what is the blackbody spectrum of the Sun? ... of the Earth?
- What is the greenhouse effect and its main causes?
- What assumptions lie behind the simplest (zero-dimensional) EBM?
- How do the key properties of the zero-dimensional EBM change when one introduces a temperature-dependent albedo?
- What is bifurcation and how does it affect the zero-dimensional EBM?
- What is the physical phenomenon described by the one-dimensional EBM but absent in the zero-dimensional EBM?
- Name the main types of energy contained in an elementary volume of ocean or atmosphere.
- What type of PDE is solved in the one-dimensional EBM?

[4]The divergence and gradient operators are given in spherical lon/lat coordinates!

2.2 Main global cycles

Cycles in the climate system

The main cycles in the Earth's climate system are

- the hydrological (water) cycle
- the carbon cycle
- the methane cycle

The hydrological cycle

The water cycle plays several major roles in the dynamics of the climate system:

- water vapor is the main greenhouse gas in the atmosphere (66 to 85% of the greenhouse effect compared to 9 to 26% for CO_2)
- water is an essential vehicle for the energy storage and transport (both due to the heat capacity of oceans and latent heat release in the atmosphere)
- water is the main ingredient in the planetary albedo changes
- is a critical factor influencing the Earth environment; its availability is essential for life and for many chemical and physical processes influencing the climate

The main reservoir of water is the Earth's crust containing ca. 10 times the amount of water in oceans. The bulk of the remaining water is contained in the oceans.

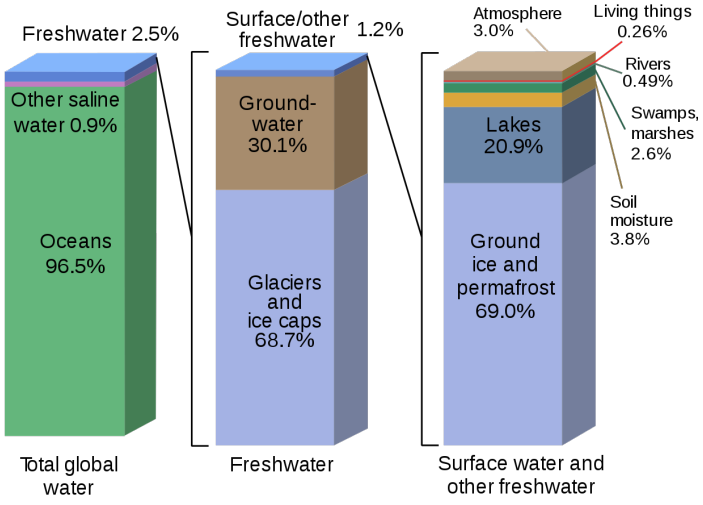


Figure 2.9: Distribution of water in the climate system (Credits: USGS).

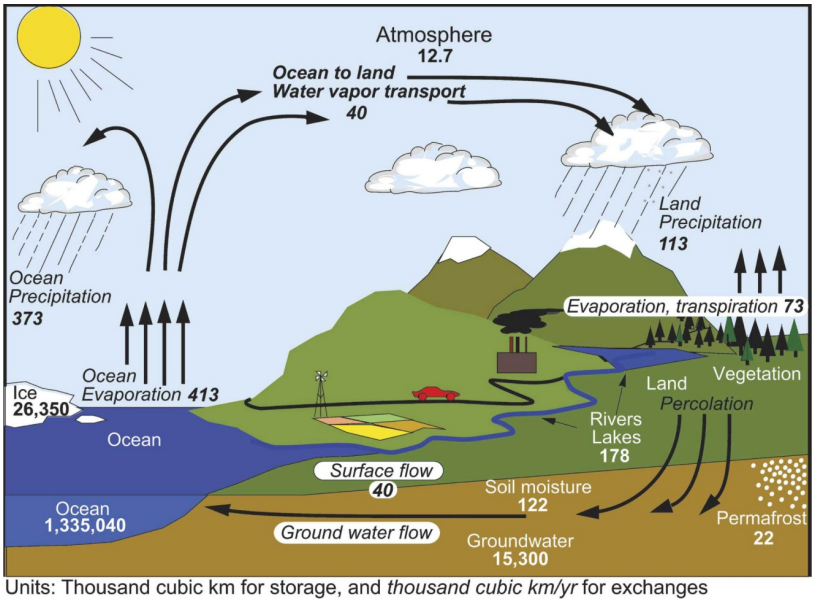


Figure 2.10: Global hydrological cycle (Trenberth et al. 2007).

The hydrological cycle over land

- Soil water content per unit volume: $\theta_{sw} = |\text{volume of water in } V|/|V|$
- Moisture content per unit surface for a layer of thickness d: $S_m = \theta_{sw} d$

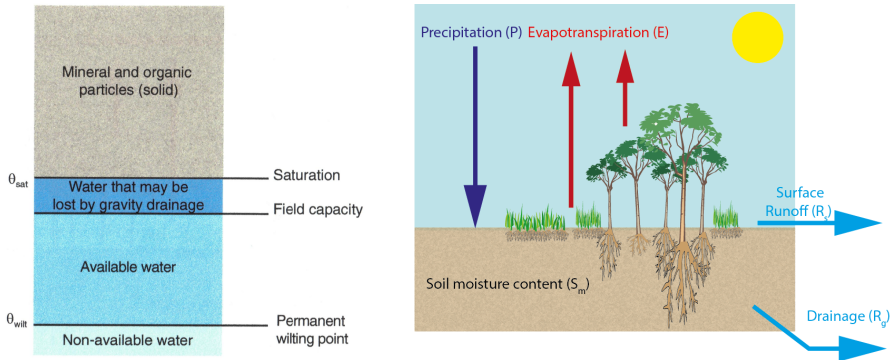


Figure 2.11: Soil moisture content (Goosse 2015).

- $\frac{dS_m}{dt} = P - E - R_s - R_g$
- P – precipitation, E – evapotranspiration, P_s – surface runoff, P_g – drainage

The hydrological cycle over oceans

Oceans compensate the $E - P$ imbalances by horizontal water transport.

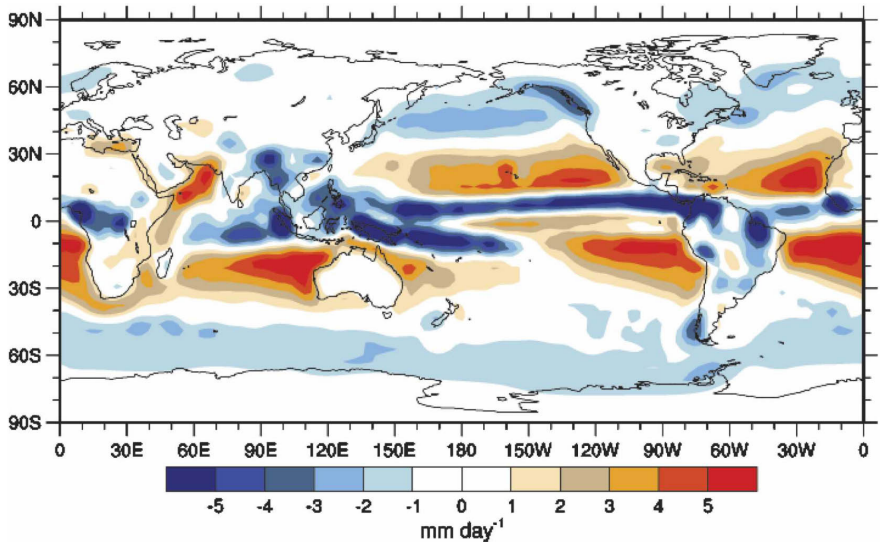


Figure 2.12: Global $E - P$ budget (Trenberth et al. 2007).

The carbon cycle

Definition 2.1. *The carbon cycle is the biogeochemical cycle of carbon exchange among the compartments of the Earth system. It describes the movement of carbon as it is recycled and reused throughout the biosphere and long-term processes of* ***carbon sequestration*** *to and release from* ***carbon sinks****.*

- As a component of two greenhouse gases, carbon dioxide (CO_2) and methane (CH_4), carbon plays a key role in the climate system
- One of the major changes brought about by human activity is the large increase in the atmospheric concentration of those two gases. The concentration of CO_2 has increased from around 280 ppm (parts per million) in 1800 to 384 ppm in 2007
- CH_4 is more reactive than CO_2 and can be oxidized to form CO_2 and H_2O. Its concentration is lower than that of CO_2, but it has increased from 725 ppb (parts per billion) to 1780 ppb in 150 years

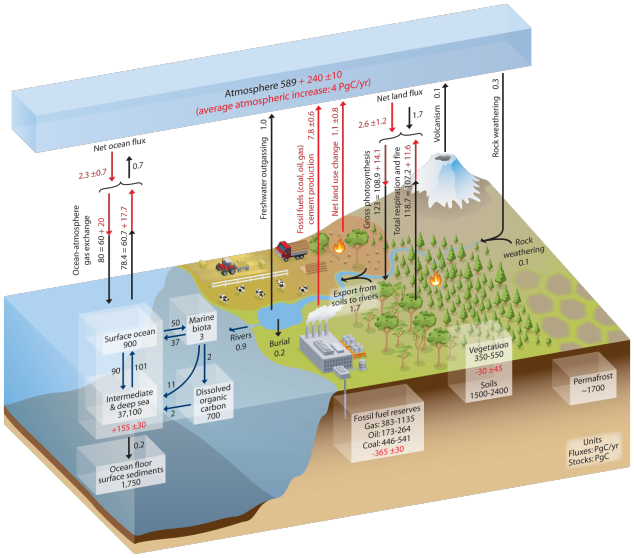


Figure 2.13: Global carbon cycle with the main natural (black) and anthropogenic (red) fluxes of carbon [PgC] (petagrams of carbon) (Ciais et al. 2013).

Oceanic carbon cycle

A flux of CO_2 between the ocean and the atmosphere Φ^{CO_2} occurs if its concentrations are at imbalance. It can be computed as

$$\Phi^{CO_2} = k^{CO_2}(p_w^{CO_2} - p_a^{CO_2}).$$

$p_w^{CO_2}, p_a^{CO_2}$ – partial pressures of CO_2 in sea water and air, k^{CO_2} – the transfer coefficient

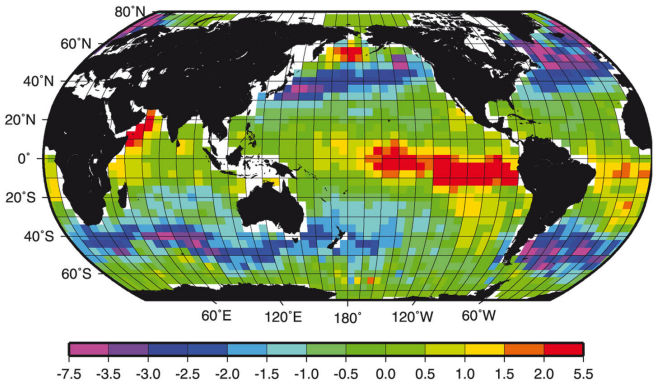


Figure 2.14: Estimates of ocean-to-atmosphere flux of CO_2 (Denman et al. 2007).

Oceanic carbon cycle: inorganic carbon

- When gaseous CO_2 is transferred from the atmosphere to the ocean, it immediately reacts with water to form carbonic acid (H_2CO_3) which dissociates, leading to the formation of bicarbonate (HCO_3^-) and carbonate ions (CO_3^{2-}):

$$\begin{aligned} CO_{2(gas)} + H_2O &\rightleftharpoons H_2CO_3 \\ H_2CO_3 &\rightleftharpoons H^+ + HCO_3^- \\ H_2CO_3^- &\rightleftharpoons H^+ + CO_3^{2-} \end{aligned}$$

- The sum of these three forms of carbon is often referred to as Dissolved Inorganic Carbon (DIC):

$$\text{DIC} = [H_2CO_3] + \left[H_2CO_3^-\right] + \left[CO_3^{2-}\right]$$

- The solubility of CO_2 strongly depends on temperature; thus CO_2 is released when the sea water circulates to higher latitudes.

Oceanic carbon cycle: organic carbon

- Next to the effect of chemical processes, the biological processes play a major role in the carbon cycle. A first important reaction is **photosynthesis** in which phytoplankton (mostly in the surface layer) use solar radiation to form organic matter from CO_2 and water:

$$6CO_2 + H_2O \rightleftharpoons C_6H_{12}O_6 + 6O_2$$

- Conversely, organic matter can be dissociated to form inorganic carbon (the reverse process of photosynthesis) by respiration and remineralization of dead phytoplankton and detritus.

- A second important biological process is related to the production of calcium carbonate (in form of calcite or aragonite) by different species, in particular to form their shells:

$$Ca^{2+} + CO_3^{2-} \rightleftharpoons CaCO_3$$

Oceanic carbon cycle: biological pumps

Definition 2.2. *The **biological pump** is the ocean's biologically driven **sequestration** of carbon from the atmosphere to the ocean interior and seafloor **sediments**.*

One distinguishes three main biological pumps

- Particles with greater density than that of sea water settle downward out of the surface layer. The net downward flux of carbon associated with this transport of organic matter is called the **soft tissue pump**
- The dissolution of calcite and aragonite mainly occurs at great depth, following the precipitation of particles and dead organisms. This leads to the downward transport of DIC called the **carbonate pump**
- A third pump, called the **solubility pump**, is associated with the sinking of cold surface water, characterized by a relatively high solubility of CO_2 and thus high DIC, to great depths at high latitudes and its resurfacing in equatorial regions with release of dissolved CO_2

Terrestrial carbon cycle

- The uptake of carbon through photosynthesis by land plants is larger than the corresponding uptake by phytoplankton, in particular in spring because of the greening of forest at mid and high latitudes and of the growth of herbaceous plants.
- About half of this primary production is directly transferred back to the atmosphere by the respiration of the land plants, and the remaining part is incorporated into leaves, wood, and roots (this fraction is called the net primary production, NPP).

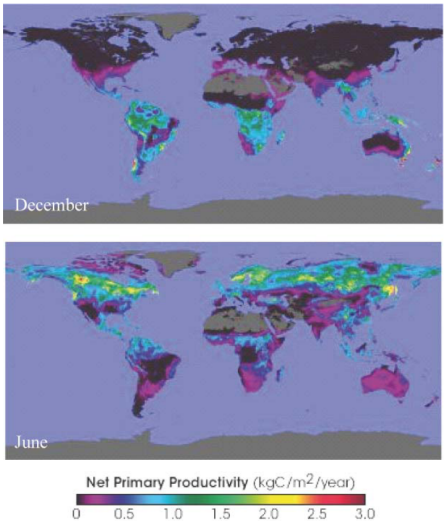


Figure 2.15: Net productivity of CO_2 over land in 2004/2005 (NASA Earth Observatory).

Geological carbon reservoirs

- The majority of the organic carbon that is exported downward from the surface layer is remineralized in the water column. In particular, the ocean is under-saturated with respect to calcite (aragonite) below 4500m (3000m) in the Atlantic and below 800 m (600m) in the Pacific. As a consequence, the long-term burial of $CaCO_3$ in the sediments to produce limestone mainly occurs in shallow seas (for instance in coral reefs).

- Averaged over the whole ocean, this long term burial corresponds to 13% of the export of $CaCO_3$ out of the surface layer. On short timescales, this is a small fraction of the whole carbon cycle, but it becomes a crucial component on timescales longer than a century. An even smaller percentage of the organic carbon is stored in the form of natural gas, oil, and coal.

- Because the sea floor spreads due to plate tectonics, sediments are transported horizontally and are eventually incorporated within the mantle through subduction along plate boundaries. At higher temperatures and pressure, limestone is transformed during subduction into calcium-silicate rocks (this is called **metamorphism**) by the reaction: The CO_2 that is released in this reaction can return to the atmosphere, in particular through volcanic eruptions.

- The plate motion also allows the calcium-silicate rocks to be uplifted to the continental surface, where they are affected by physical and chemical **weathering**.

- The products of this reaction are transported by rivers to the sea where they can compensate for the net export of $CaCO_3$ by sedimentation.

Overall sedimentation, subduction, metamorphism, and weathering form a closed loop that takes place over millions of years and is sometimes referred to as the long term inorganic carbon cycle.

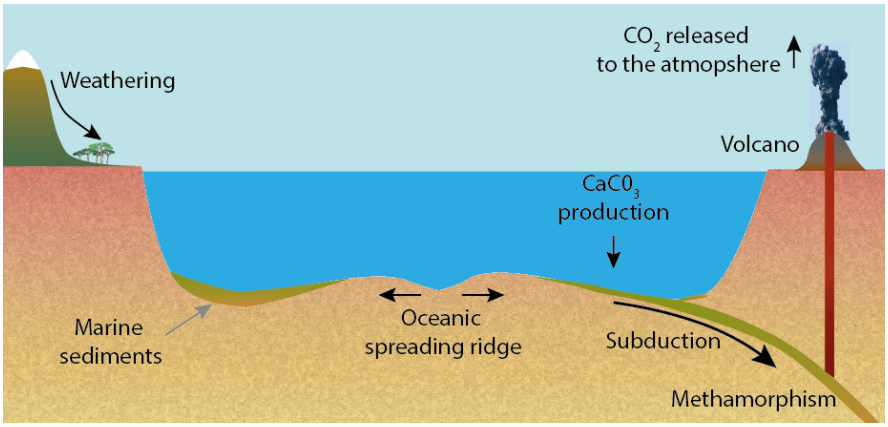


Figure 2.16: Long term inorganic carbon cycle through sedimentation, subduction, metamorphism, and weathering (Goosse 2015).

The methane cycle

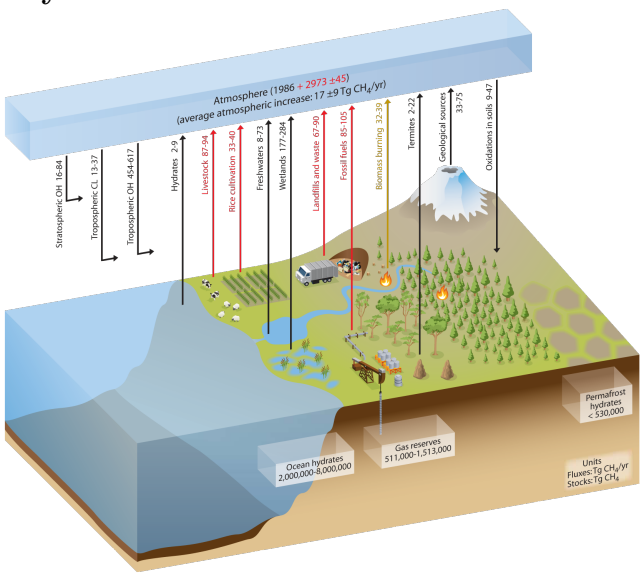


Figure 2.17: Global methane cycle with the main natural (black) and antropogenic (red) fluxes of CH_4 Tg(CH_4) (teragrams of CH_4) (Ciais et al. 2013).

Questions

- What part of the Earth system contains the largest reserves of water?

- Explain the idea of 'available water' in the soil.
- What is the carbon cycle and why is it important for climate modeling?
- Describe the main aspects of the inorganic carbon cycle in the ocean.
- What are the two main effects of biological processes on the ocean carbon cycle.
- What is a bilogical pump? Describe the main biological pumps in the ocean.
- What mechanisms affect the carbon budget in geological carbon reservoirs?

2.3 Modeling biogeochemistry and carbon cycle

Biogeochemistry modeling for ocean

The most common type of model used in ocean biogeochemistry is a coupled set of advection-diffusion-reaction equations for **tracer** concentrations c_i

$$\partial_t c_i + \nabla \cdot (\boldsymbol{v} c_i - K \nabla c_i) = \sum_{k=1}^{n} R(c_i, c_k) + \Psi_{atm} + \Psi_{sed}$$

- $R(c_i, c_k)$ – reaction rate between c_i and c_k
- Ψ_{atm} – flux of c_i from/to atmosphere
- Ψ_{sed} – flux of c_i from/to sea bed

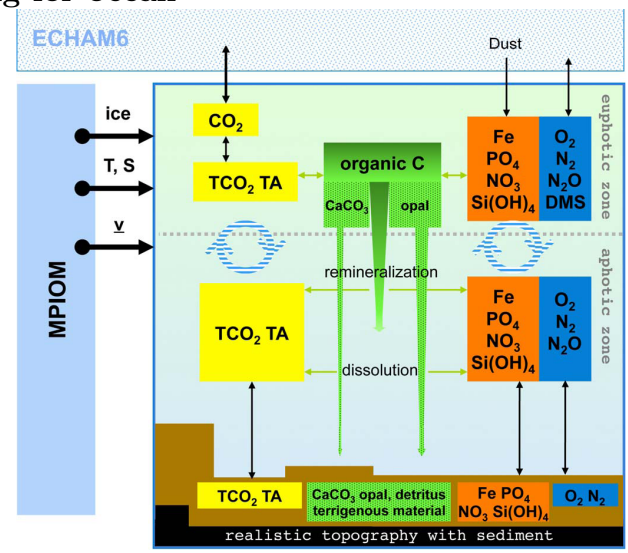


Figure 2.18: Schematic of ocean biogeochemistry model HAMOCC (Ilyina et al. 2013).

Tracers and reactions

The tracers can be various chemical, biological, or physical quantities

- DIC, alkalinity

- dissolved gases (O_2, CO_2, ...) and isotopes (e.g. ^{14}C)
- nutrients (Fe, PO_4, NO_3, ...)
- phytoplankton or various **functional groups** thereof (functional groups are distinguished by their size, nutrient requirement, and their biogeochemical role in the ecosystem but not necessary by their taxa)
- zooplankton or various functional groups thereof
- detritus (sinking particulate organic matter)

The advective $\nabla\cdot(\boldsymbol{v}c_i)$ and the diffusive $-\nabla\cdot(K\nabla c_i)$ transport are prescribed by the oceanic equations of motion; the key part of the biogeochemistry model are the reaction and source/sink terms. Therefore, the model can be generally studied using an ODE system given by

$$\frac{dc_i}{dt} = \sum_{k=1}^{n} R(c_i, c_k) + \Psi_{atm} + \Psi_{sed}, \qquad i = 1, \ldots, n.$$

Population models: Exponential growth

The equations of this type are called **population models** and are widely used in population dynamics modeling.

The simplest model of this type considers the population P of one species with unlimited resources and net growth proportional to the population size

$$\frac{dP}{dt} = kP, \qquad P(0) = P_0.$$

Its solution can be obtained analytically

$$P(t) = P_0 e^{kt}.$$

It implies an exponential growth for $k > 0$ or an exponential decay for $k < 0$.

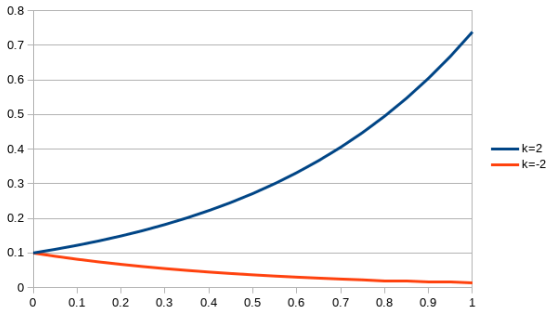


Figure 2.19: Exponential growth with P_0=0.1.

Population models: The logistic model
Exponential growth is usually unrealistic: No resources are infinite.
The one-species model can be augmented by introducing the maximum population M called the 'carrying capacity'

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{M}\right), \qquad P(0) = P_0.$$

Its solution can also be obtained analytically

$$P(t) = \frac{P_0 M}{P_0 + (M - P_0)e^{-kt}}$$

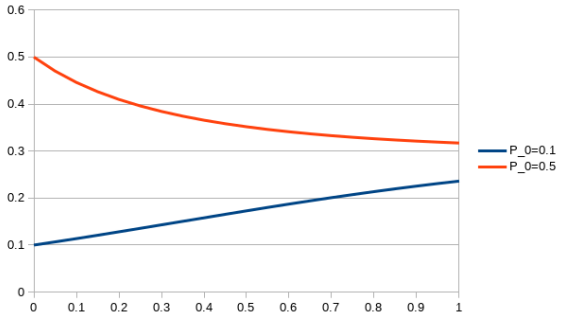


Figure 2.20: Logistic model with $M = 0.3$.

Population models: Predator-prey model
Predator-Prey model (also known as Lotka-Volterra equations) considers two species R (predator) and B (prey) that dynamically interact with each other. The prey species is assumed to have unlimited food source.

$$\frac{dB}{dt} = \mu_B B - \alpha BR$$
$$\frac{dR}{dt} = -\mu_R R + \beta BR$$

$\mu_B, \mu_R, \alpha, \beta$ are positive constants. $\mu_B B$ is the 'natural' growth rate of the prey species, $-\mu_R R$ is the 'natural' death rate of the predator species, $-\alpha BR$ is the death rate of prey due to predator (proportional to the meet frequency BR), βBR is the growth rate of predator due to the availability of prey.

This system is difficult to solve analytically, but one can find equilibria of the model (points with $\frac{dB}{dt} = \frac{dR}{dt} = 0$) and plot the solution trajectories in the phase space.

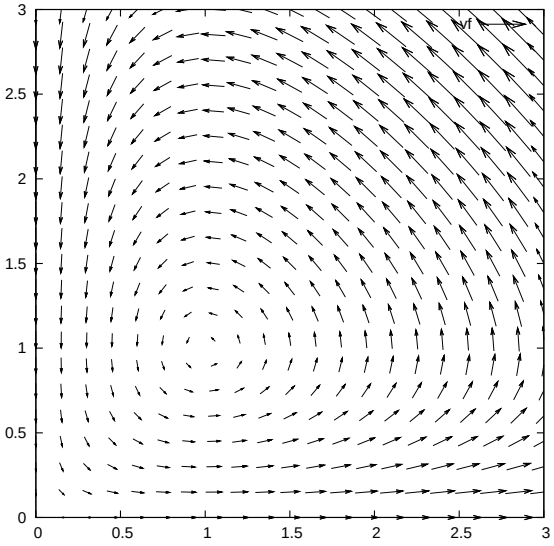


Figure 2.21: Phase-space plot of the predator-prey model. Prey population (horizontal axis), predator population (vertical axis).

Ecosystem modeling: NPZ and NPZD models

NPZ The standard approach to the full ecosystem modeling usually considers at least three compartments: (N)utrients, (P)hytoplankton, and (Z)ooplankton

NPZD For long-term ocean biogeochemistry, carbon cycle, and climate studies, one also considers the forth compartment: (D)etrius

For more detailed modeling, each compartment can be subdivided into as many subcompartments as needed: Current state-of-the-art models consider 40-70 different compartments in total.
Phytoplankton and zooplankton can, in particular, be modeled by identifying the main functional groups as tracers and representing the exchange processes between them. For phytoplankton such functional groups can include: Small phytoplankton (nano- and picoplankton), large (nonsilicifying) phytoplankton, silicifying phytoplankton, calcifying phytoplankton, N_2-fixing phytoplankton, etc.

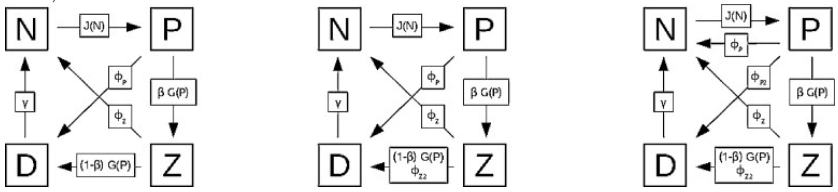


Figure 2.22: Schematic of three different NPZD models (Heinle, Slawig 2013).

$$\frac{dN}{dt} = -J(N,I)P + \Phi_Z Z + \gamma D$$
$$\frac{dP}{dt} = \left(J(N,I) - \Phi_P\right) P - G(\epsilon, g, P)Z$$
$$\frac{dZ}{dt} = \left(\beta G(\epsilon, g, P) - \Phi_Z\right) Z$$
$$\frac{dD}{dt} = \Phi_P P + \left((1-\beta)G(\epsilon, g, P) + \Phi_Z\right) Z - \gamma D$$

Symbol	Definition
I	light intensity
$J(\cdot,\cdot)$	phytoplankton growth rate
γ	detritus remineralization rate
$\Phi_{P,Z}$	mortality rates
β	assimilation efficiency of zooplankton
$G(\cdot,\cdot,\cdot)$	zooplankton grazing rate
ϵ	grazing encounter rate
g	maximum grazing rate

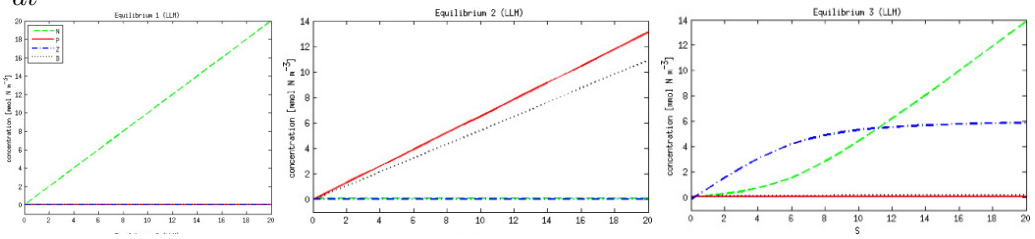


Figure 2.23: Three equilibria of the NPZD model depending on the initial mass S distributed among N, P, Z, and D compartments (Heinle, Slawig 2013).

- The first equilibrium corresponds to the situation when no phytoplankton growth occurs (e.g. not enough sunlight) and is given by $N = N_0, P = Z = D = 0$
- In the second equilibrium, the phytoplankton reaches its growth limit on the available nutrients in the absence of zooplankton
- In the third equilibrium, all compartments are present, zooplankton reaches its growth limit using up the available phytoplankton

Environmental study with an NPZ model

- Study site: Galveston Bay (Texas)
- Study length: 90 days
- Initial conditions: localized sources of phyto- and zooplankton, no nutrients
- Boundary conditions: constant river inflow of nutrients

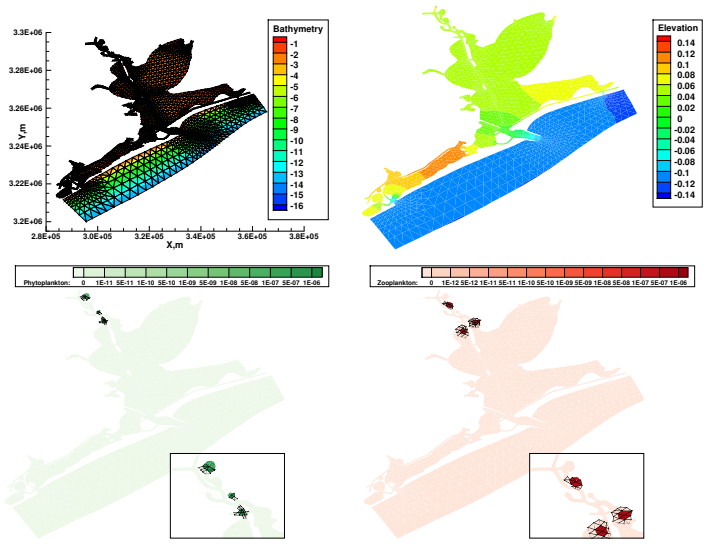


Figure 2.24: Model setup and initial conditions (Hajduk et al. 2018).

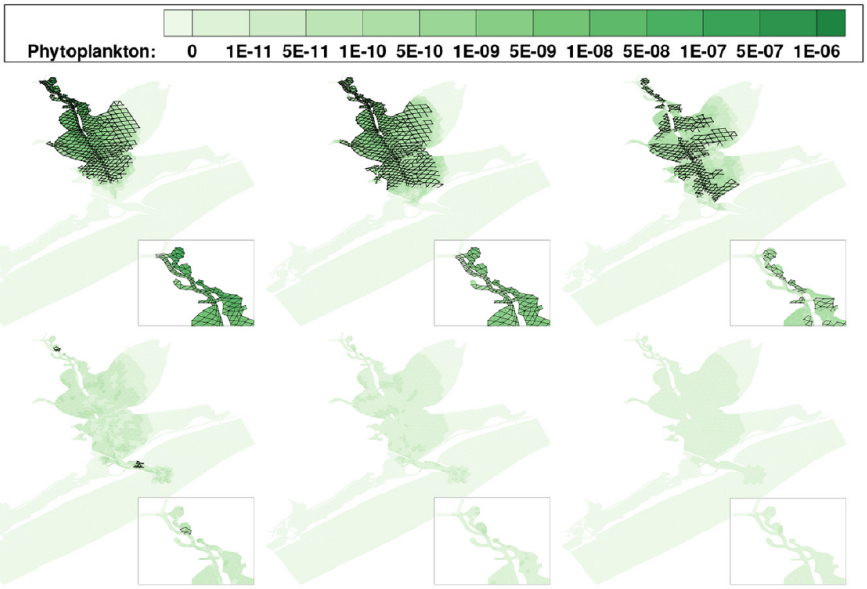


Figure 2.25: Development of phytoplankton over 90 days (Hajduk et al. 2018).

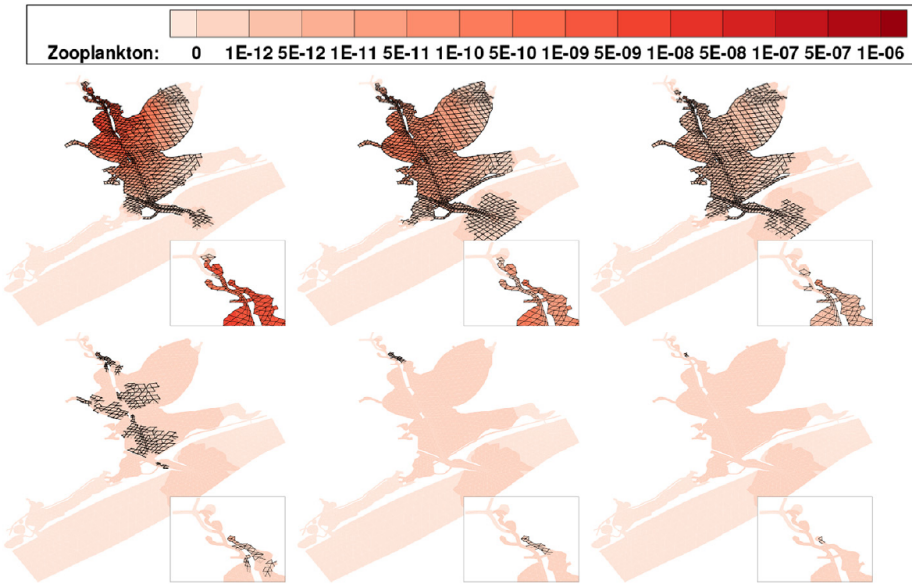


Figure 2.26: Development of zooplankton over 90 days (Hajduk et al. 2018).

Questions

- What type of equation system is used to model ocean biogeochemistry?
- Can this model be run in a standalone mode (without the ocean dynamics model)?
- What is a 'tracer'? What quantities can be tracers?
- What is the simplest population model? Why is it not a realistic one?
- Describe briefly the main features of a predator-prey model.
- What is an equilibrium of a predator-prey model? How can it be characterized mathematically?
- What is the main idea behind the NPZ and NPZD models?
- How can the NPZ and NPZD models be made more detailed?

2.4 Geophysical flows

Specifics of large-scale geophysical flows

Two main ingredients distinguish the geophysical fluid dynamics (GFD) from the traditional fluid mechanics: The effects of rotation and of stratification. The relative influence of either one leads to peculiarities exhibited only by geophysical flows. Also the scales of motion are several orders of magnitude greater than those of a typical engineering application.

- Earth's rotation gives rise to two additional accelerations entering the momentum equations as forcing terms: The **Coriolis** and **centrifugal** forces. The former plays a crucial role in geophysical flows (both, oceanic and atmospheric), whereas the latter is usually neglected
- **Stratification** arises because naturally occurring flows involve fluids of different densities (e.g., warm and cold air, fresh and saline water). Gravity forces the heavier fluid to go down and the lighter one to go up resulting, under equilibrium conditions, in a **stably stratified** fluid consisting of vertically stacked horizontal layers. Fluid motions act to disturb this equilibrium, whereas gravity strives to restore it

Scales of motion

Remark 2.3. *To discern whether a physical process is dynamically important in any particular situation, one introduces so-called* **scales of motion**. *These are dimensional quantities expressing the overall magnitude of the variables under consideration and understood as estimates rather than precise values. A specific choice of temporal or spatial scales requires a clear understanding of the physics of the problem and can be ambiguous for phenomena taking place on variety of different scales; e.g., the atmosphere exhibits variations on daily (weather) and decadal (climate) scales. The selection of scales then reflects the choice of processes being investigated.*

- The key scales are time (T), length (L), and velocity (U)
- For flows, in which the density plays an important role, additional scales are introduced: Mean density (ρ_0), typical range of density variations ($\Delta\rho$), height (H) over which such density variations occur

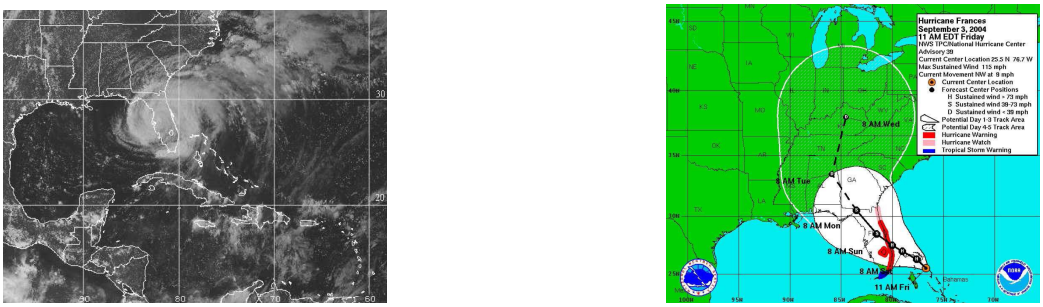


Figure 2.27: Hurricane Frances over Florida on 5 September 2004. Satellite image (left), model prediction made on 3 September (right) (NOAA 2004).

Hurricane Frances (2004) during her course over the southeastern USA (Fig. 2.27) had a nearly circular form spanning ca. 7.5° (830 km) with sustained surface winds from 59 to 69 m/s. These data suggest the following choice of scales: L = 800 km, U = 60 m/s, and T = 2×10^5 s (= 55.6 h). This fits well to the typical hurricane tracks displaying appreciable change in direction and speed of propagation over 2-day intervals.

Importance of rotation

A major effect of the Coriolis force is to impart certain vertical flow patterns to the fluid. In rapidly rotating, homogeneous fluids, this effect can be so strong that the flow displays strict columnar motions; that is, all particles along the same vertical line evolve in concert and retain their vertical alignment over long periods of time.

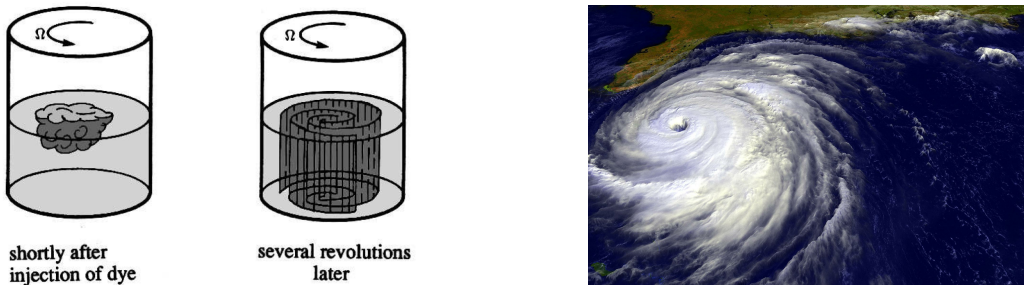


Figure 2.28: Left: An initially amorphous cloud of dye is transformed after several rotations of the vessel into vertical sheets called Taylor curtains (Cushman-Roisin, Beckers 2011). Right: Hurricane Frances approaching Florida (NASA 2004).

Definition 2.4. *The Earth's rotation rate is given by* $\Omega = \frac{2\pi}{\textit{length of } \textbf{sidereal day}}$, *where sidereal day (ca. 23 hours 56 minutes) is the shortest period of time for a fixed star to be seen at the same position from the same point on Earth. Sidereal day takes into account the periods of rotation of Earth around its axis and around the Sun.*

We consider the effects of rotations to be important, if either

- time scale of motion T is comparable to or longer than the period of Earth's rotation, i.e. $T \geq 2\pi/\Omega \quad \Leftrightarrow \quad \frac{2\pi}{\Omega T} \leq 1$
- or particle traveling at velocity U covers distance L or less during the period of Earth's rotation, i.e.

$$U\, 2\pi/\Omega \leq L \quad \Leftrightarrow \quad \varepsilon = \frac{2\pi U}{\Omega L} \leq 1$$

The following table gives an intuitive idea of lengths and velocities for which rotational effects start to matter:

L	1m	10m	100m	1km	10km	100km	1000km	6371km (Earth radius)
U $\leq$	$0.012\frac{mm}{s}$	$0.12\frac{mm}{s}$	$1.2\frac{mm}{s}$	$1.2\frac{cm}{s}$	$12\frac{cm}{s}$	$1.2\frac{m}{s}$	$12\frac{m}{s}$	$74\frac{m}{s}$

Obviously, in most engineering applications the above conditions are not met. This includes, e.g.,

- the flow of water at a speed of 5 m/s in a turbine 1 m in diameter giving $\varepsilon \approx 4 \times 10^5$
- or the air flow past a 5-m wing on an airplane flying at 100 m/s resulting in $\varepsilon \approx 2 \times 10^6$

On the other hand, an ocean current flowing at 10 cm/s a distance of 100 km or a wind blowing at 10 m/s in a 1000-km-wide anticyclonic formation do meet the criteria for the rotational effects to be important.

Importance of stratification

- Geophysical fluids typically consist of fluid masses of different densities, which under constant gravitational action tend to arrange themselves in vertical columns corresponding to a state of minimal potential energy. However, motions continuously disturb this equilibrium raising dense fluid and sinking light fluid. The corresponding increase of potential energy happens at the expense of kinetic energy thereby slowing the flow
- Also the opposite situation happens: Previously disturbed stratification returns toward equilibrium, potential energy converts into kinetic energy, and the flow gains momentum
- Therefore the dynamical importance of stratification can be evaluated by comparing potential and kinetic energies

Introducing σ, the ratio of kinetic energy (given by $\frac{1}{2}\rho_o U^2$) to the change in potential energy (given by $\Delta\rho g H$ and equal to the energy needed to vertically exchange a parcel of fluid with density $\rho_0 + \Delta\rho$ with a parcel of fluid with density ρ_0 over height H), we obtain

$$\sigma = \frac{\frac{1}{2}\rho_o U^2}{\Delta\rho g H},$$

where $g = 9.81$ m/s^2 is the average gravitational acceleration on Earth. Depending on the value of σ, three standard situations can arise:

- For $\sigma \sim 1$, a typical change in potential energy consumes a sizable chunk of the kinetic energy, therefore stratification is important
- For $\sigma \ll 1$, the kinetic energy level is insufficient to substantially disturb the stratification, thus the fluid is mostly at rest
- Finally, $\sigma \gg 1$ means that stratification hardly affects the flow

Combining rotation with stratification
Of special interest is the case when rotation and stratification effects are equally important, yet neither dominates over the other: $\varepsilon \sim 1$ and $\sigma \sim 1$.

- Omitting the constant factors 2π and $1/2$, we arrive at

$$L \sim \frac{U}{\Omega} \quad \text{and} \quad U \sim \sqrt{\frac{\Delta\rho}{\rho_0} g H}.$$

- Combining the above estimates yields the fundamental length scale

$$L \sim \frac{1}{\Omega}\sqrt{\frac{\Delta\rho}{\rho_0} g H},$$

which for the Earth conditions ($\Omega = 7.29 \times 10^{-5}$ s^{-1}, $g = 9.81$ m/s^2) takes the following values:

  - atmosphere ($\rho_0 = 1.2$ kg/m^3, $\Delta\rho = 0.03$ kg/m^3, $H = 5000$ m)
    $L \sim 500$ km, $\quad U \sim 30$ m/s $\quad \Rightarrow$ size/wind speed of weather patterns
  - ocean ($\rho_0 = 1028$ kg/m^3, $\Delta\rho = 2$ kg/m^3, $H = 1000$ m)
    $L \sim 60$ km, $\quad U \sim 4$ m/s $\quad \Rightarrow$ width/speed of major ocean currents

The Coriolis force

Let the X- and Y-axes form the inertial framework of reference and the x- and y-axes have the same origin but rotate counterclockwise at angular rate Ω. The corresponding unit vectors are denoted $(\boldsymbol{I}, \boldsymbol{J})$ and $(\boldsymbol{i}, \boldsymbol{j})$. At time t, the rotating x-axis makes an angle Ωt with the fixed X-axis giving

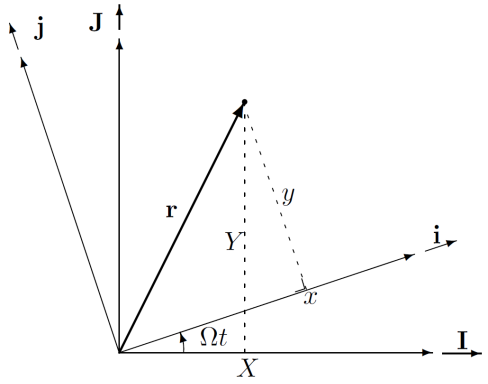


Figure 2.29: Fixed (X, Y) and rotating (x, y) frames of reference.

$$\begin{aligned}
\boldsymbol{i} &= \boldsymbol{I}\cos(\Omega t) + \boldsymbol{J}\sin(\Omega t),\\
\boldsymbol{j} &= -\boldsymbol{I}\sin(\Omega t) + \boldsymbol{J}\cos(\Omega t),\\
\boldsymbol{I} &= \boldsymbol{i}\cos(\Omega t) - \boldsymbol{j}\sin(\Omega t),\\
\boldsymbol{J} &= \boldsymbol{i}\sin(\Omega t) + \boldsymbol{j}\cos(\Omega t).
\end{aligned} \tag{2.2}$$

For a point $\boldsymbol{r} = X\boldsymbol{I} + Y\boldsymbol{J} = x\boldsymbol{i} + y\boldsymbol{j}$, one obtains

$$\begin{aligned}
x &= X\cos(\Omega t) + Y\sin(\Omega t),\\
y &= -X\sin(\Omega t) + Y\cos(\Omega t).
\end{aligned} \tag{2.3}$$

Differentiating (2.3) with respect to time yields

$$\begin{aligned}\frac{dx}{dt} &= \frac{dX}{dt}\cos(\Omega t) + \frac{dY}{dt}\sin(\Omega t)\underbrace{-\Omega X\sin(\Omega t) + \Omega Y\cos(\Omega t)}_{=\Omega y},\\ \frac{dy}{dt} &= -\frac{dX}{dt}\sin(\Omega t) + \frac{dY}{dt}\cos(\Omega t)\underbrace{-\Omega X\cos(\Omega t) - \Omega Y\sin(\Omega t)}_{=-\Omega x}.\end{aligned} \tag{2.4}$$

Writing out relative velocity $\boldsymbol{u}$ using the coordinate vectors of the rotating frame gives

$$\boldsymbol{u} = \frac{dx}{dt}\boldsymbol{i} + \frac{dy}{dt}\boldsymbol{j} =: u\boldsymbol{i} + v\boldsymbol{j}.$$

A similar expression holds for absolute velocity vector $\boldsymbol{U}$ in the inertial frame

$$\boldsymbol{U} = \frac{dX}{dt}\boldsymbol{I} + \frac{dY}{dt}\boldsymbol{J}.$$

Next, we express the absolute velocity vector $\boldsymbol{U}$ via the coordinate vectors of the rotating frame using transformations for $\boldsymbol{I}, \boldsymbol{J}$ from (2.2)

$$\begin{aligned}\boldsymbol{U} &= \left(\frac{dX}{dt}\cos(\Omega t) + \frac{dY}{dt}\sin(\Omega t)\right)\boldsymbol{i} + \left(-\frac{dX}{dt}\sin(\Omega t) + \frac{dY}{dt}\cos(\Omega t)\right)\boldsymbol{j}\\ &=: U\boldsymbol{i} + V\boldsymbol{j}.\end{aligned}$$

Combining (2.4) with the above equation results in

$$U = u - \Omega y, \qquad V = v + \Omega x. \tag{2.5}$$

These equalities state that the absolute velocity is equal to the relative velocity plus the correction velocity arising from the rotation of the reference framework.
Differentiating once more with respect to time obtains

$$\begin{aligned}\tfrac{d^2x}{dt^2} &= \tfrac{d^2X}{dt^2}\cos(\Omega t) + \tfrac{d^2Y}{dt^2}\sin(\Omega t) + 2\,\Omega\underbrace{\left(-\tfrac{dX}{dt}\sin(\Omega t) + \tfrac{dY}{dt}\cos(\Omega t)\right)}_{=V}\\ &\quad -\Omega^2\underbrace{\left(X\cos(\Omega t) + Y\sin(\Omega t)\right)}_{=x},\\ \tfrac{d^2y}{dt^2} &= -\tfrac{d^2X}{dt^2}\sin(\Omega t) + \tfrac{d^2Y}{dt^2}\cos(\Omega t) - 2\,\Omega\underbrace{\left(\tfrac{dX}{dt}\cos(\Omega t) + \Omega\tfrac{dY}{dt}\sin(\Omega t)\right)}_{=U}\\ &\quad -\Omega^2\underbrace{\left(-X\sin(\Omega t) + Y\cos(\Omega t)\right)}_{=y}.\end{aligned} \tag{2.6}$$

Similarly to the velocity, we obtain accelerations in each frame of reference

$$\begin{aligned}
\boldsymbol{a} &= \tfrac{d^2x}{dt^2}\boldsymbol{i} + \tfrac{d^2y}{dt^2}\boldsymbol{j} =: a\boldsymbol{i} + b\boldsymbol{j},\\
\boldsymbol{A} &= \tfrac{d^2X}{dt^2}\boldsymbol{I} + \tfrac{d^2Y}{dt^2}\boldsymbol{J}\\
&= \left(\tfrac{d^2X}{dt^2}\cos(\Omega t) + \tfrac{d^2Y}{dt^2}\sin(\Omega t)\right)\boldsymbol{i} + \left(-\tfrac{d^2X}{dt^2}\sin(\Omega t) + \tfrac{d^2Y}{dt^2}\cos(\Omega t)\right)\boldsymbol{j}\\
&=: A\boldsymbol{i} + B\boldsymbol{j}.
\end{aligned}$$

Using (2.5) we can write (2.6) compactly as

$$\begin{aligned}
a &= A + 2\,\Omega V - \Omega^2 x, \qquad & b &= B - 2\,\Omega U - \Omega^2 y,\\
A = a &\underbrace{-\ 2\,\Omega v}_{\substack{\text{Coriolis}\\ \text{acceleration}}} \underbrace{-\ \Omega^2 x}_{\substack{\text{Centrifugal}\\ \text{acceleration}}}, & B = b &\underbrace{+\ 2\,\Omega u}_{\substack{\text{Coriolis}\\ \text{acceleration}}} \underbrace{-\ \Omega^2 y}_{\substack{\text{Centrifugal}\\ \text{acceleration}}}.
\end{aligned}$$

Introducing the vector rotation $\boldsymbol{\Omega} = \Omega\boldsymbol{k}$, where $\boldsymbol{k}$ is the unit vector corresponding to the vertical coordinate common for both, inertial and rotating systems, the same result can be written in a vector form

$$\begin{aligned}
\boldsymbol{U} &= \boldsymbol{u} + \boldsymbol{\Omega}\times\boldsymbol{r},\\
\boldsymbol{A} &= \boldsymbol{a} + \underbrace{2\,\boldsymbol{\Omega}\times\boldsymbol{u}}_{\substack{\text{Coriolis}\\ \text{acceleration}}} + \underbrace{\boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\boldsymbol{r})}_{\substack{\text{Centrifugal}\\ \text{acceleration}}},
\end{aligned}$$

where '$\times$' denotes the vector product.

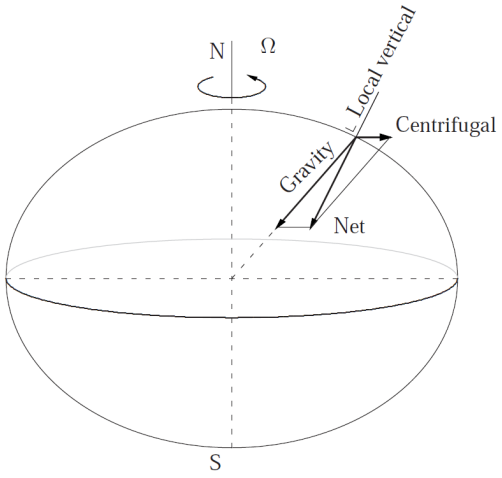


Figure 2.30: Illustration of the centrifugal force and of flattening of the rotating Earth (Cushman-Roisin, Beckers 2011).

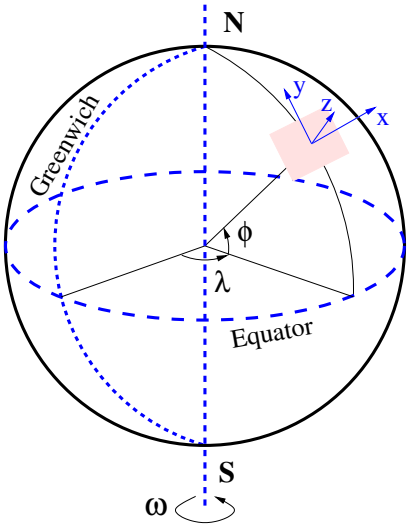


Figure 2.31: Definition of the local coordinate system: x-axis eastward, y-axis northward, z-axis upward.

- The centrifugal force is directed outward, perpendicularly to the axis of rotation; combined with the **gravitation** of the Earth, it constitutes Earth's **gravity** field. Earth's shape (called **geoid**) corresponds to an **equipotential** surface of this field

- The direction of the the gravity force resulting from the vector addition of the gravitational and centrifugal forces (Fig. 2.30) is thus perpendicular to the surface of the Earth and coincides with the local vertical

- In the 'traditional' local Cartesian coordinate system, the x-axis is directed eastward, the y-axis northward, and the z-axis upward (see Fig. 2.31). In this framework, Earth's rotation is expressed as

$$\boldsymbol{\Omega} = \Omega \cos(\varphi)\, \boldsymbol{j} + \Omega \sin(\varphi)\, \boldsymbol{k}$$

- Using notation $\boldsymbol{u} = (u, v, w)^T$, the Coriolis term has the form

$$2\,\boldsymbol{\Omega} \times \boldsymbol{u} = 2\,\Omega \begin{pmatrix} \cos(\varphi) w - \sin(\varphi) v \\ \sin(\varphi) u \\ -\cos(\varphi) u \end{pmatrix}$$

Questions

- What distinguishes geophysical flow systems from engineering ones?
- What is the idea behind the scales of motion and what scales are commonly used in connection with geophysical flows?
- What is the sidereal day and how is it related to Earth's rotation rate?
- For which types of flow are the rotation effects important?
- What is stratification and how is it related to the potential energy of the fluid?
- How is it possible to determine the importance of stratification for a given flow system?
- What role does the Coriolis force play in the momentum equations?
- What is the difference between the form of the Coriolis term for a rotating disk and that in the geographic coordinates?
- Why does the gravity force generally point into a different direction than the gravitational force?

2.5 Basics of Continuum Mechanics

Brief introduction to Continuum Mechanics

With big thanks to Prof. Ian Hewitt

Idea of a continuum: Although fluids and solids consist of molecules, the continuum approximation treats the material as having a continuous distribution of mass. This assumption is only valid on spatial and temporal scales that are much larger than the largest molecular scales.

Each 'point' of the material is ascribed properties such as density, velocity, pressure, or temperature. Some of them are related to each other by constitutive laws – essentially by empirical parameterizations of the unresolved molecular mechanics of the material.

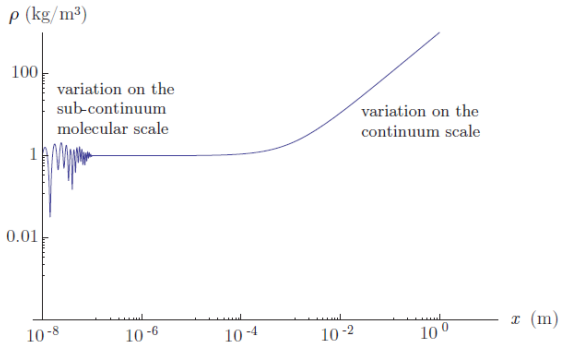


Figure 2.32: Density at different scales (Powers 2020).

Eulerian and Lagrangian coordinates

Two different coordinate systems are used in continuum mechanics.

- In **Eulerian coordinates** $(\boldsymbol{x}, t)$, spatial coordinate $\boldsymbol{x}$ is fixed in space (that is, in a fixed reference frame). A parcel of fluid will generally move through different coordinates as time t evolves
- In **Lagrangian coordinates** $(\boldsymbol{X}, t)$, spatial coordinate $\boldsymbol{X}$ is fixed in the material; it labels the same parcel of fluid for all time. Commonly $\boldsymbol{X}$ is chosen as the Eulerian coordinate of the parcel at the initial time

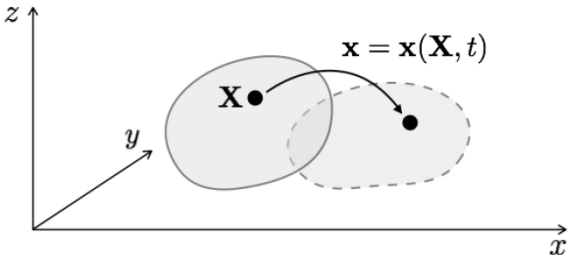


Figure 2.33: Parcel of fluid with Lagrangian coordinates $\boldsymbol{X}$ follows Eulerian path $\boldsymbol{x}(\boldsymbol{X}, t)$.

The velocity of this parcel of fluid with respect to Eulerian coordinates is given by

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

The material derivative

- Consider a function $f(\boldsymbol{x}, t)$ described in Eulerian coordinates.

- The partial derivative with respect to time $\dfrac{\partial f}{\partial t}(\boldsymbol{x},t)$ gives the rate of change of f at point $\boldsymbol{x}$ and time t.
- Alternatively, one can use a coordinate transformation to express f in Lagrangian coordinates $\widehat{f}(\boldsymbol{X},t) \coloneqq f(\boldsymbol{x}(\boldsymbol{X},t),t)$ and calculate the time derivative of $\widehat{f}$ using the chain rule of differentiation

$$\frac{\partial \widehat{f}}{\partial t} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\underbrace{\frac{\partial x}{\partial t}}_{=u} + \frac{\partial f}{\partial y}\underbrace{\frac{\partial y}{\partial t}}_{=v} + \frac{\partial f}{\partial z}\underbrace{\frac{\partial z}{\partial t}}_{=w} = \frac{\partial f}{\partial t} + \boldsymbol{u}\cdot\nabla f =: \frac{Df}{Dt}.$$

- The time derivative $\dfrac{Df}{Dt}(\boldsymbol{x},t)$ describes the rate of change of f for a fluid parcel moving through point $\boldsymbol{x}$ at time t. It is called the **material derivative** (also known as the total or advective derivative).

Mass conservation: Eulerian coordinates
Consider an arbitrary fixed volume V within the fluid of density $\rho(\boldsymbol{x},t)$ moving with the velocity $\boldsymbol{u}(\boldsymbol{x},t)$. The mass within this volume can only change due to the movement across its boundary ∂V resulting in the (integral) statement of mass conservation

$$\frac{d}{dt}\int_V \rho(\boldsymbol{x},t)d\boldsymbol{x} = -\int_{\partial V}\rho(\boldsymbol{x},t)\boldsymbol{u}(\boldsymbol{x},t)\cdot\boldsymbol{n}\,ds,$$

where $\boldsymbol{n}$ denotes the outward unit normal to the boundary of V.
Noting that V is not time-dependent we can pull the time derivative into the first integral. Assuming that $\rho\boldsymbol{u}$ is differentiable in space we obtain after applying the Gauss theorem to the surface integral

$$\int_V \left(\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{u})\right) d\boldsymbol{x} \;=\; 0.$$

Since V is arbitrary, the differential statement of mass conservation holds

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{u}) = 0.$$

Mass conservation: Lagrangian coordinates
The statement of mass conservation for a material point is then given by

$$\frac{D\rho}{Dt} + \rho\nabla\cdot\boldsymbol{u} = 0.$$

If the fluid is **incompressible**, the density of a fluid parcel does not change, thus

$$\frac{D\rho}{Dt} = \frac{\partial \rho}{\partial t} + \boldsymbol{u} \cdot \nabla \rho = 0$$

implying that the fluid is divergence-free

$$\nabla \cdot \boldsymbol{u} = 0.$$

The latter equation is called the **continuity** equation.

Questions

- What is the main assumption behind the continuum approach?
- What are Eulerian coordinates, and what is the meaning of the time derivative in these coordinates?
- What are Lagrangian coordinates, and hat is the meaning of the time derivative in these coordinates? How is it called?
- What fluid is called incompressible?

Forces in continuum mechanics

Continuum mechanics deals with deformable as opposed to rigid bodies.

Definition 2.5.
- *A solid is a deformable body that possesses* **shear** *strength, i.e. a solid resists shear deformations (deformations parallel to the material surface on which they act)*
- *Fluids, on the other hand, do not resist shear deformation*

In the classical dynamics of Newton and Euler, the motion of a material body is caused by the action of externally applied forces which are assumed to be of two kinds:

- **Body forces**
- **Surface forces**

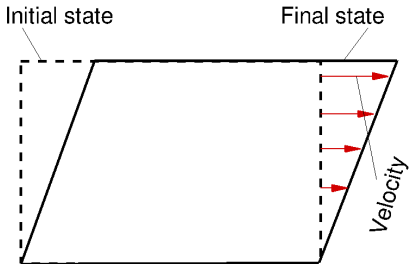


Figure 2.34: Shear deformation schematic.

Body forces

- Body forces act on all parts of the body in the same way; they usually arise from an interaction of an external field (e.g. gravity or electromagnetic) with the body
- Examples include the gravity, electric, or magnetic forces
- Also the Coriolis force is counted among the body forces

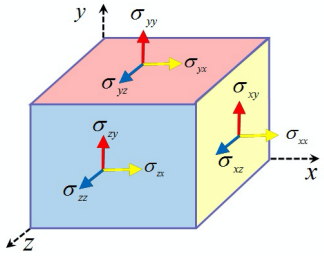
Surface forces

- Surface (or contact) forces arise due to actions of or interactions between the molecules of the material (e.g. attraction or collision); they can be present throughout the body and on its external surface
- Their specific magnitude and direction depends on the **shape and orientation** of the surface being considered
- Examples include pressure or friction
- Surface forces usually manifest themselves in the form of **stress**

Mechanical stress

- In continuum mechanics, stress is a physical quantity that expresses the internal forces that neighboring particles of a continuous material exert on each other.
- Stress σ is defined as the force F applied to a material divided by the material's cross-sectional area A: $\sigma = F/A$.
- The stress state in a material is described by means of the **Cauchy stress tensor** (matrix) $\boldsymbol{\sigma}$ whose components represent the force per unit area in the jth direction on a surface with a normal pointing in the ith direction.

$$\boldsymbol{\sigma} = \sigma_{ij} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix}$$



The Cauchy stress tensor is symmetric $\sigma_{ij} = \sigma_{ji}$, therefore one needs only 6 components of $\boldsymbol{\sigma}$ to fully describe the stress state at a material point.
The Cauchy stress tensor allows to calculate the stress for an arbitrarily oriented surface, thus, for a surface with unit normal vector $\boldsymbol{n}$, it is given by

$$\boldsymbol{s} = \boldsymbol{\sigma} \cdot \boldsymbol{n} \quad \text{or} \quad s_i = \sum_{j=1}^{3} \sigma_{ij} n_j, \ i = 1, 2, 3.$$

We define the pressure p to be the negative mean of the diagonal components of the stress tensor

$$p = -\frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33})$$

and then decompose the stress tensor into

$$\boldsymbol{\sigma} = -p\boldsymbol{I} + \boldsymbol{\tau},$$

where $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{\tau}$ is the **deviatoric (viscous) stress tensor**. For non-viscous fluids (e.g. an ideal gas), one has $\boldsymbol{\tau} = 0$.

Momentum conservation

Consider once again an arbitrary fixed volume V within the fluid of density $\rho(\boldsymbol{x}, t)$ moving with the velocity $\boldsymbol{u}(\boldsymbol{x}, t) = (u, v, w)^T$.

Its momentum is given by $\displaystyle\int_V \rho\, \boldsymbol{u}\, d\boldsymbol{x}$.

Changes of momentum can occur due to the movement of fluid into and out of the volume and due to the action of forces resulting in the (integral) statement of momentum conservation

$$\frac{d}{dt}\int_V \rho\, \boldsymbol{u}\, d\boldsymbol{x} = -\int_{\partial V} \rho\, \boldsymbol{u}\, (\boldsymbol{u} \cdot \boldsymbol{n})\, ds - \int_V \rho\, (g\boldsymbol{k} + 2\boldsymbol{\Omega} \times \boldsymbol{u})\, d\boldsymbol{x} + \int_{\partial V} \boldsymbol{\sigma} \cdot \boldsymbol{n}\, ds,$$

where g is the gravity acceleration, and $\boldsymbol{k} = (0, 0, 1)^T$.
Using the Gauss theorem and the fact that V is time-independent and arbitrary, we arrive at the differential statement of momentum conservation

$$\frac{\partial(\rho\, \boldsymbol{u})}{\partial t} + \nabla \cdot (\rho \boldsymbol{u} \otimes \boldsymbol{u}) + \rho\, g\boldsymbol{k} + 2\rho\, \boldsymbol{\Omega} \times \boldsymbol{u} - \nabla \cdot \boldsymbol{\sigma} = 0,$$

where $\otimes$ is the tensor (outer) product defined as $\boldsymbol{u} \otimes \boldsymbol{u} = \begin{pmatrix} u^2 & uv & uw \\ uv & v^2 & vw \\ uw & vw & w^2 \end{pmatrix}$.

Navier–Stokes equations
The equations of mass and momentum conservation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0, \tag{2.7}$$

$$\frac{\partial (\rho\, u)}{\partial t} + \nabla \cdot (\rho\, u\, \boldsymbol{u}) + 2\rho\, \Omega\, (\cos(\varphi)\, w - \sin(\varphi) v) \; - \nabla \cdot (\sigma_{xx}, \sigma_{xy}, \sigma_{xz}) = 0, \tag{2.8}$$

$$\frac{\partial (\rho\, v)}{\partial t} + \nabla \cdot (\rho\, v\, \boldsymbol{u}) + 2\rho\, \Omega \sin(\varphi)\, u - \nabla \cdot (\sigma_{yx}, \sigma_{yy}, \sigma_{yz}) = 0, \tag{2.9}$$

$$\frac{\partial (\rho\, w)}{\partial t} + \nabla \cdot (\rho\, w\, \boldsymbol{u}) + \rho\, g - 2\rho\, \Omega \cos(\varphi)\, u - \nabla \cdot (\sigma_{zx}, \sigma_{zy}, \sigma_{zz}) = 0 \tag{2.10}$$

are called the **Navier–Stokes equations**; they represent the fundamental model in fluid mechanics used to simulate nearly any type of fluid motion. The primary unknowns in (2.7)–(2.10) are density ρ and velocity $\boldsymbol{u} = (u, v, w)$. However, in order to obtain a closed system, one also needs a constitutive expression for $\boldsymbol{\sigma}$ relating it to the primary unknowns. Such expressions rely on the knowledge of the **rheology** of the material.

Questions

- What type of deformation is called a shear deformation?
- How is a solid (fluid) defined in continuum mechanics?
- What kinds of forces are considered in the classical mechanics of Newton and Euler?
  - Describe the physical mechanism and give examples of body forces.
  - Describe the physical mechanism and give examples of surface forces.
  - Why pressure is a surface force?
- What is the mechanical stress and why do we need a tensor to describe it mathematically? How is it calculated on a given surface?
- How do we define pressure using the Cauchy stress tensor? Why do we call it isotropic?

- What is the deviatoric stress tensor and why is it called 'viscous'?
- What types of conservation relationships are described by the Navier–Stokes equations?
- What are the primary unknowns of the Navier–Stokes equations?

Rheology

- Rheology is the science of deformation and flow. Knowing rheological properties of a material allows us to provide an empirical relationship between the deformation or velocity variations and the stress.
- For fluids (as opposed to solids), the stress can be expressed in terms of velocity gradients. The pressure p represents the **isotropic** (i.e. acting equally in all directions) part of the stress tensor, whereas the deviatoric (viscous) stress tensor $\boldsymbol{\tau}$ represents deviations from this isotropic state.
- For **Newtonian fluids** such as water or air, viscous stresses depend linearly on velocity gradients.

Our representation of the stress tensor $\boldsymbol{\sigma}$ can be then expressed as

$$\boldsymbol{\sigma} = -p\boldsymbol{I} + \boldsymbol{\tau} = -p\boldsymbol{I} - \frac{2}{3}\mu(\nabla \cdot \boldsymbol{u})\boldsymbol{I} + \mu\left(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T\right), \quad \nabla\boldsymbol{u} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{pmatrix},$$

where μ is called the dynamic viscosity coefficient.

Boundary conditions

- In order to have a unique solution (i.e. to be well-posed), the system (2.7)–(2.10) must be complemented with boundary conditions that correctly represent physical processes taking place at the exterior boundaries.
- The standard PDE boundary conditions such as Dirichlet or Neumann are building blocks for more elaborate expressions needed in fluid dynamics.
- The boundary conditions for the Navier–Stokes system are of either of the two types: **kinematic** (i.e., describing the movement at the boundary) or **dynamic** (i.e., describing the forces applied to the boundary surface).

- The specific boundary conditions prescribed vary according to the type of the boundary.

Boundary conditions: sea bed

- The standard kinematic boundary condition for a viscous fluid at a rigid boundary (wall) is **'no slip'**: $\boldsymbol{u} = 0$.
- If the fluid is inviscid, this changes to **'no normal flow'** $\boldsymbol{u} \cdot \boldsymbol{n} = 0$, where $\boldsymbol{n}$ denotes a normal to the boundary surface.
- Let the 2D function $z = z_b(x, y)$ describe a (time-independent) sea bed (ocean) or a land surface (atmosphere) boundary. The 'no normal flow' equation can then be put into a 2D form

$$\boldsymbol{n} = \nabla(z - z_b) = \left(-\frac{\partial z_b}{\partial x}, -\frac{\partial z_b}{\partial y}, 1\right)^T \Rightarrow 0 = \boldsymbol{u}_{|_{z=z_b}} \cdot \boldsymbol{n} = u_{|_{z=z_b}} \frac{\partial z_b}{\partial x} + v_{|_{z=z_b}} \frac{\partial z_b}{\partial y} - w_{|_{z=z_b}}.$$

- Alternatively, one can take the material derivative of $z - z_b$ and take into account that this boundary is not moving, and that no material crosses it

$$\frac{D}{Dt}(z - z_b) = 0 \Leftrightarrow \underbrace{\frac{\partial(z - z_b)}{\partial t}}_{0} + u_{|_{z=z_b}} \underbrace{\frac{\partial(z - z_b)}{\partial x}}_{-\frac{\partial z_b}{\partial x}} + v_{|_{z=z_b}} \underbrace{\frac{\partial(z - z_b)}{\partial y}}_{-\frac{\partial z_b}{\partial y}} + w_{|_{z=z_b}} \underbrace{\frac{\partial(z - z_b)}{\partial z}}_{1} = 0.$$

Boundary conditions: free surface

- In a similar manner, namely using the material derivative, one can derive the kinematic boundary condition at the oceanic free surface. We define the vertical coordinate of the free surface with respect to some zero-level (e.g. the mean sea level) by the 2D function $z = \xi(x, y, t)$ and note that the free surface elevation is generally a time-dependent function
- Similarly to the bottom boundary condition, we take into account that no material crosses this boundary and take the material derivative of $z - \xi$

$$\frac{D}{Dt}(z-\xi)=0 \quad \Leftrightarrow \quad \underbrace{\frac{\partial(z-\xi)}{\partial t}}_{-\frac{\partial \xi}{\partial t}} + u_{|_{z=\xi}} \underbrace{\frac{\partial(z-\xi)}{\partial x}}_{-\frac{\partial \xi}{\partial x}} + v_{|_{z=\xi}} \underbrace{\frac{\partial(z-\xi)}{\partial y}}_{-\frac{\partial \xi}{\partial y}} + w_{|_{z=\xi}} \underbrace{\frac{\partial(z-\xi)}{\partial z}}_{1} = 0.$$

- This results in the equation that can be used to determine the free surface elevation

$$\frac{\partial \xi}{\partial t} + u_{|_{z=\xi}} \frac{\partial \xi}{\partial x} + v_{|_{z=\xi}} \frac{\partial \xi}{\partial y} - w_{|_{z=\xi}} = 0.$$

Boundary conditions: dynamic boundary conditions
Whereas kinematic boundary conditions prescribe the movement, the dynamic ones describe forces acting on the surface (usually in the form of stress). Thus at the air-sea interface (free surface of the ocean), one usually postulates the continuity of pressure

$$p_{|_{z=\xi}} = p_a,$$

where p_a denotes the atmospheric pressure, and prescribes the wind stress

$$\boldsymbol{\tau}_{|_{z=\xi}} \cdot \boldsymbol{n} = \boldsymbol{\tau}_s,$$

where $\boldsymbol{\tau}_s$ is the surface wind stress.
Similar conditions can be used to model friction effects at the sea bed

$$\boldsymbol{\tau}_{|_{z=z_b}} \cdot \boldsymbol{n} = \boldsymbol{\tau}_b,$$

where $\boldsymbol{\tau}_b$ denotes the bottom stress.

Questions

- What type of relationship is given with the help of rheology?
- What type of fluid is called Newtonian?
- What kinds of standard boundary conditions are used in connection with the Navier–Stokes equations?
- What type of kinematic boundary condition is prescribed at the sea bed?
- How do we derive the kinematic boundary condition at the free surface?
- What do we prescribe as dynamic boundary conditions at the free surface? sea bed?

2.6 Ocean

2.6.1 Ocean physics

Ocean: Composition and physical properties

- The ocean covers ca. 71% of Earth's surface
- It has an average depth of about 3700 m
- Sea water is composed of 96.5% water, the remaining 3.5% are dissolved salts, particles, gases, and organic matter
- The salt (NaCl) represents about 85% of the dissolved material; its content is usually measured in [ppt] (parts per thousand) or [psu] (practical salinity units) – the latter is based on the electrical conductivity of the sea water; both measurements usually give very similar results
- Sea water is slightly compressible (density increases with pressure); the sea water density also increases with salinity and decreases with increasing temperature
- Salinity also influences the freezing point that varies between 0°C for pure water and -1.8°C for water with salinity of 35 psu

Ocean: Circulation types

- The driving forces of ocean circulation are Earth gravity and rotation, wind, gravity of other celestial bodies (mainly the Sun and the Moon causing tides), and tectonic events
- Large-scale ocean studies are mainly concerned with two distinct types of circulation: Wind-induced surface currents and the so-called **thermohaline** circulation caused by the interplay between variations in sea water density and winds, Earth rotation, and topography
- Coastal ocean dynamics is also strongly forced by tides and weather systems, which – due to their relatively short time scales – play only a secondary role in the large-scale dynamics.

Ocean: Wind-induced circulation

- The **surface** ocean circulation is mainly driven by the winds
- At mid-latitudes, the atmospheric **westerlies** induce eastward currents in the ocean while the **trade winds** are responsible for westward currents in the tropics

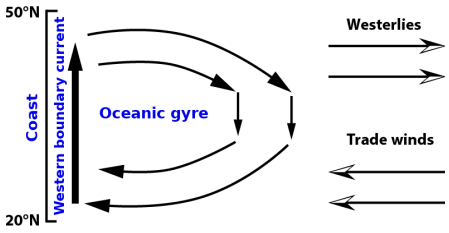


Figure 2.35: Oceanic gyres & western boundary currents (Goosse 2015).

- Because of continental barriers, those currents form loops called the subtropical **gyres** characterized by an intensification along the western boundaries of the oceans (eastern coasts of continents). These include well-known strong currents such as the **Gulf Stream** off the east coast of the USA and the **Kuroshio** off Japan
- These currents generally run parallel to the winds, whereas the equatorial countercurrents run in the direction opposite to the trade winds

In the Southern Ocean, in the absence of continental barriers the Antarctic Circumpolar Current (ACC) connects all ocean basins. This is one of the strongest currents on Earth transporting about $0.13 \times 10^9 \mathrm{m}^3\mathrm{s}^{-1}$ of water.

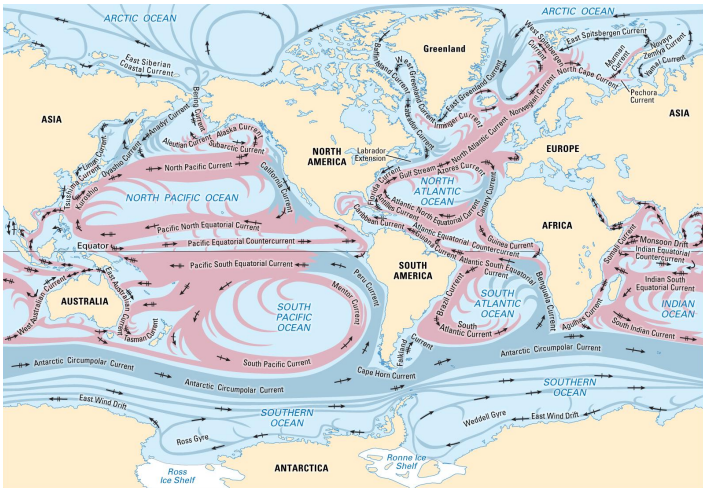


Figure 2.36: Map of important surface currents (Encyclopedia Britannica).

- Because of Earth's rotation, the wind-induced ocean transport is perpendicular to the wind stress (to the right in the Northern, to the left in the Southern Hemisphere – cf. Fig. 2.37). This transport called the Ekman transport plays an important role in explaining the path of the wind-driven surface currents (Fig. 2.36)

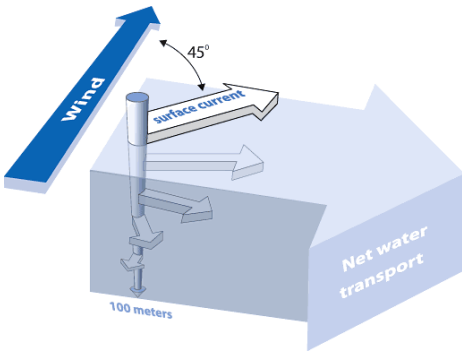


Figure 2.37: Ekman transport schematic for the Northern Hemisphere (NOAA).

- Along coastlines or if the transport has horizontal variations, this can lead to surface convergence/divergence that has to be compensated by vertical movements in the ocean

- The Ekman transport direction to the north in the Northern and to the south in the Southern Hemisphere results in a divergence at the equator that has to be compensated by an upwelling there (c.f. Fig. 2.38 (left)). This phenomenon is called the **equatorial upwelling**

- For winds blowing parallel to coast, a similar mechanism leads to a **coastal upwelling** compensating for the transport away from the coast (Fig. 2.38 (right))

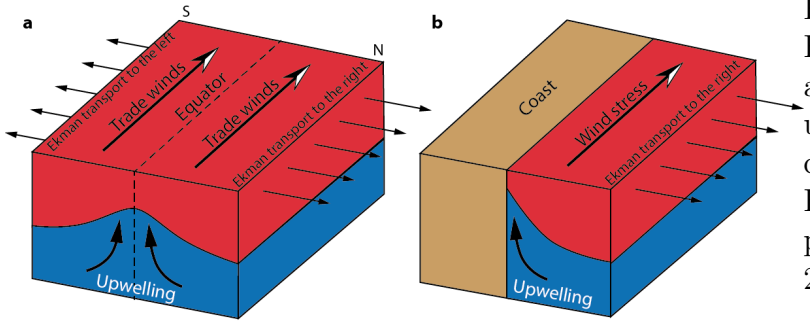


Figure 2.38: Equatorial (left) and coastal upwelling (right) caused by the Ekman transport (Goosse 2015).

Ocean: Thermohaline circulation

- At high latitudes, low temperatures and relatively high salinity cause the surface water to sink to great depths. This process, often referred to as **deep oceanic convection**, is only possible in a few places in the world, mainly in the North Atlantic and in the Southern Ocean
- These water masses (called the **deep water**) are then slowly transported from the North Atlantic southward along the western boundary of the Atlantic towards the Southern Ocean. From there, it is transported to the other oceanic basins where it slowly surfaces
- The return flow to the sinking regions is achieved through surface and intermediate depth circulation
- The thermohaline circulation is quite slow. The time needed for water masses formed in the North Atlantic to reach the Southern Ocean is of the order of a century. If the whole cycle is taken into account, the time scale is estimated as several centuries to a few millennia depending of the exact location and mechanism studied

- The thermohaline circulation (Fig. 2.39), which is associated with currents at all depths, is often called the **global conveyor belt**; it is driven by density gradients, but winds also play a significant role
- This circulation transports huge amounts of water, salts, energy, and nutrients that play an important role in the global climate dynamics.

Ocean: Salinity

- The sea surface salinity is strongly influenced by the freshwater fluxes at the surface reaching a maximum in subtropical areas because of the large evaporation and low rainfall there. The high precipitation rates induce lower salinity at the equator, while the weak evaporation is responsible for the lower salinity observed at mid and high latitudes

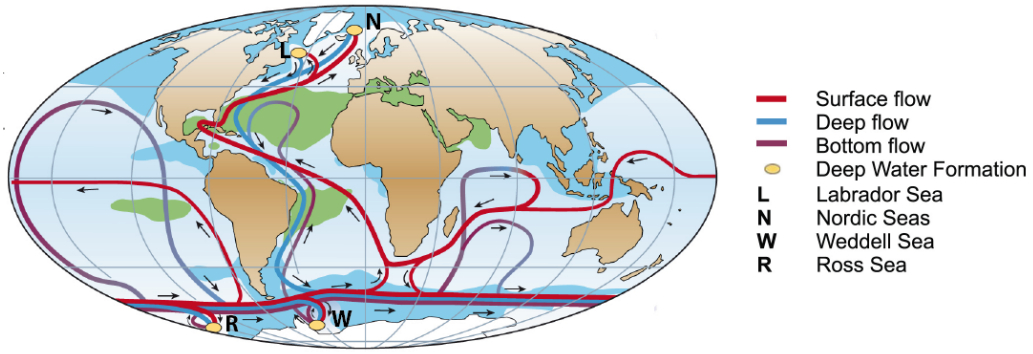


Figure 2.39: Schematic of the oceanic thermohaline circulation (Ramstorf 2002).

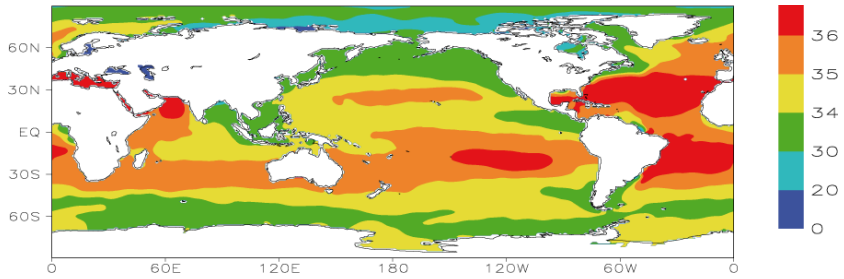


Figure 2.40: Annual mean sea surface salinity [psu] (Levitus 1998).

- River input also has a large regional impact with low values close to the mouths of the Amazon and Mississippi rivers

Ocean: Temperature at the surface

- Because of the strong interactions between the ocean and the atmosphere, the sea surface temperature (SST) is very close to the temperature of the air above it. One exception is the polar regions where sea ice insulates the ocean from the cold polar atmosphere

- The uppermost tens of meters of ocean water at mid- and high latitudes show a clear seasonal cycle

- When temperature rises in spring and summer, a shallow layer (generally less than 40 m) warms up and stabilizes the water column. The

water below this layer is insulated from the surface and thus conserves the properties acquired by contact with the air in winter

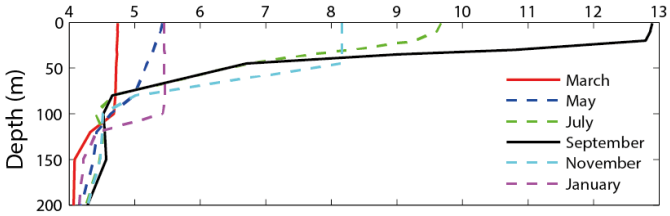


Figure 2.41: Monthly mean temperature profiles at a mid-latitude site in the Northern Hemisphere (PMEL/NOAA).

Ocean: Mixed layer

In winter, stirring by the winds and cooling at the surface tend to destabilize the water column and generate shallow convection inducing thereby strong mixing. This homogenizes a surface layer called the **oceanic mixed layer** (Fig. 2.42) that has a typical winter depth of 50 to 100 m but may reach several hundred meters in some regions

This process gives rise to a region with strong vertical variations in density under the summer extent of the mixed layer that is called the **seasonal thermocline**

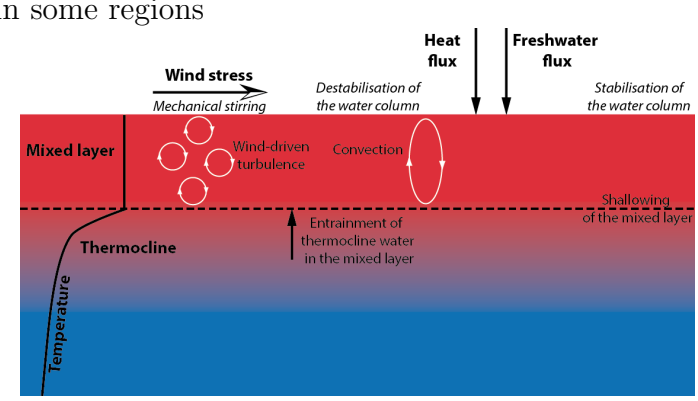


Figure 2.42: Oceanic mixed layer (Goosse 2015).

Ocean: Deep water

- Below the mixed layer (except in polar regions), a temperature gradient is observed causing the **permanent thermocline** (Fig. 2.43)
- This shows that the majority of the ocean has strong stratification meaning that light water sits above dense water as required by the vertical stability of the water column
- In the deep ocean, the vertical density variations are much weaker

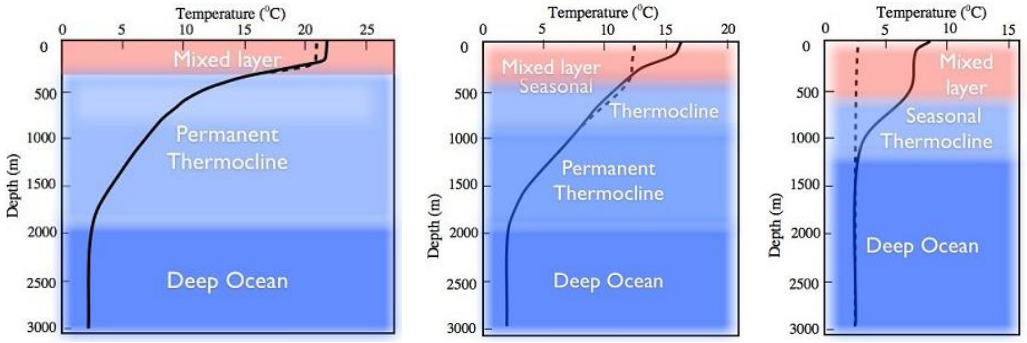


Figure 2.43: Thermocline: Tropics (left), midlatitudes (center), and polar (right).

- The temperature and salinity of sea water are modified by interactions with the atmosphere only in the oceanic mixed layer
- Thus **water mass** formation and transformation mainly occur close to the surface. When these waters flow beneath the mixed layer, they tend to keep the properties they have acquired close to the surface
- As a consequence, the path of important water masses such as North Atlantic Deep Water (NADW), Antarctic Bottom Water (AABW), or Antarctic Intermediate Water (AAIW) can easily be followed from their region of formation on temperature vertical sections (Fig 2.44)

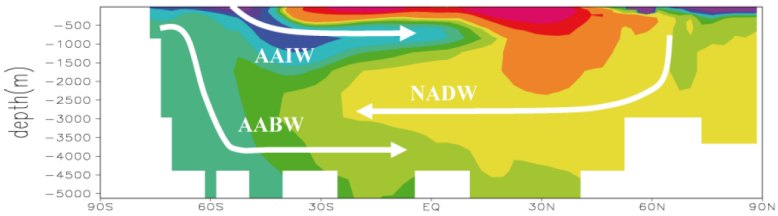


Figure 2.44: Zonally averaged temperature in Atlantic Ocean (Levitus 1998).

Questions

- What types of circulation are relevant for the global ocean dynamics?
- What is a subtropical gyre and how does it arise?

- What is the mechanism behind the Ekman transport?
- What is the deep oceanic convection? Where can it happen?
- What is the typical time scale of the thermohaline circulation?
- What is the global conveyor belt, and why is it called thus?
- What factors strongly affect the global distribution of ocean salinity?
- What is the oceanic mixed layer and what important processes are taking place in it?
- What is the seasonal thermocline, and where is it **not** present?
- What is the permanent thermocline and how does it influence the interaction between the surface and deep water in different regions?
- How do you interpret the term 'water mass', and what is the role of water masses in the global eco-system?

2.6.2 Ocean circulation modeling

Ocean: The equation of state
The description of the fluid system is not complete until a relation between density and pressure called the **equation of state** is specified

- For pure water at ordinary pressures and temperatures, the statement can be as simple as $\rho = \text{const}$. In the ocean, however, water density is a complicated function of pressure, temperature, and salinity
- For most applications, it can be assumed that the density of seawater is independent of pressure (incompressibility) and linearly dependent on both temperature and salinity according to

$$\rho(T, S) = \rho_0(1 - \alpha(T - T_0) + \beta(S - S_0)),$$

where T is the temperature and S the salinity. ρ_0, T_0, and S_0 are the reference values of density, temperature, and salinity, respectively, whereas α and β are the coefficients of thermal expansion and saline 'contraction'. Typical seawater values are $\rho_0 = 1028\ \text{kg/m}^3$, $T_0 = 10°\text{C}$, $S_0 = 35$ psu, $\alpha = 1.7 \times 10^{-4}\ \text{K}^{-1}$, and $\beta = 7.6 \times 10^{-4}$

Ocean: The hierarchy of circulation models

The starting point for the derivation of the model hierarchy for ocean (Fig. 2.45) are the mass and momentum conservation (Navier–Stokes) equations in a frame of reference rotating with angular velocity $\boldsymbol{\Omega}$.

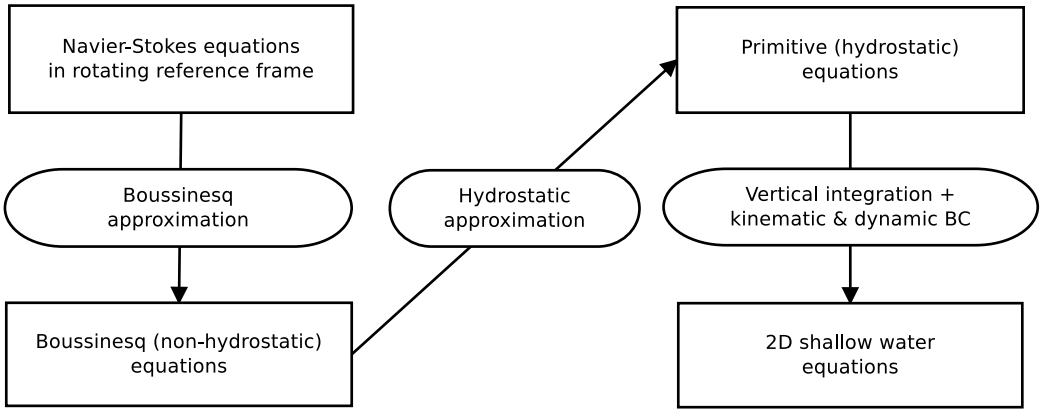


Figure 2.45: Hierarchy of ocean circulation models.

Ocean: The Navier–Stokes equations

The Navier–Stokes equations are the main mathematical model describing the motion of **viscous** fluids.

Denoting by $\boldsymbol{\tau}$ the viscous stress tensor and by p the pressure, the conservative (also called divergence) form of the Navier-Stokes system is

$$\overbrace{\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u})}^{\text{Conservation of mass}} = 0, \tag{2.11}$$

$$\overbrace{\underbrace{\frac{\partial (\rho \boldsymbol{u})}{\partial t}}_{\substack{\text{change in} \\ \text{momentum}}} + \underbrace{\nabla \cdot (\rho \boldsymbol{u} \otimes \boldsymbol{u})}_{\substack{\text{momentum} \\ \text{advection}}} + \underbrace{\nabla p + \rho\, g\, \boldsymbol{k}}_{\substack{\text{pressure and} \\ \text{gravity forces}}} + \underbrace{2\, \rho\, \boldsymbol{\Omega} \times \boldsymbol{u}}_{\substack{\text{Coriolis} \\ \text{force}}} - \underbrace{\nabla \cdot \boldsymbol{\tau}}_{\substack{\text{viscous} \\ \text{forces}}}}^{\text{Conservation of momentum}} = 0, \tag{2.12}$$

where $\otimes$ is the notation for the tensor (outer) product.

In continuum mechanics, the Navier–Stokes system is the foundation for a range of key mathematical models. The standard special cases are:

- Neglecting viscosity effects ($\boldsymbol{\tau} = \text{const}$) results in the **Euler equations** widely used (usually in combination with an energy conservation equation) for modeling compressible fluids (atmospheric flows, acoustics, airspace applications)

- Assuming constant density ($\rho = \rho_0$) results in the **incompressible** Navier-Stokes equations that pose a **saddle point** problem

$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u}) + \frac{\nabla p}{\rho_0} + g\,\boldsymbol{k} + 2\,\boldsymbol{\Omega} \times \boldsymbol{u} - \frac{\nabla \cdot \boldsymbol{\tau}}{\rho_0} = 0, \tag{2.13}$$

$$\nabla \cdot \boldsymbol{u} = 0. \tag{2.14}$$

- Neglecting, in addition, the advection term results in the **Stokes system** – a standard model for very viscous (and frequently also non-Newtonian) flows; in climate modeling, this system and its simplifications are used to model sea and land ice flows

In the ocean modeling, the mass and momentum conservation equations must be complemented by the temperature and salinity transport equations given as advection-diffusion equations with μ_r denoting the diffusivity coefficient for the corresponding constituent

$$\frac{\partial r}{\partial t} + \nabla \cdot (\boldsymbol{u}\, r) - \nabla \cdot (\mu_r\, \nabla r) = 0, \qquad r \in \{T, S\} \tag{2.15}$$

and an equation of state for density

$$\rho = \rho(T, S, p). \tag{2.16}$$

Conservation of internal energy is expressed by the transport equation for temperature; the exchange between the mechanical and internal energy is usually not modeled explicitly since its effects are comparably minor for the majority of ocean applications.

Questions

- What physical relationship is expressed by an equation of state for sea water?
- What is the starting point for the derivation of the ocean circulation models?
- What physical phenomena are accounted for in the Navier-Stokes equations?
- In which situations is it possible to neglect viscous effects?
- How is the Navier-Stokes system modified to model incompressible fluids?
- What is the difference between Stokes and Navier-Stokes systems?
- What additional relationships – aside of the Navier-Stokes equations – are needed in ocean circulation models?

Ocean: The Boussinesq (non-hydrostatic) system

- The density-driven circulation plays a major role in ocean systems whenever large temporal scales are important. Motions caused by density gradients due to varying salinity and temperature are generally slow since the local variations in the sea water density in the world oceans rarely exceed 0.5%
- This fact motivates the fundamental assumption made in nearly every ocean model, the so-called

Boussinesq approximation

Replace the actual water density with its reference value ρ_0 everywhere except in the gravity forcing term $\rho g \boldsymbol{k}$ in the momentum equation (2.12).

Applying this approximation to system (2.12)–(2.11), dividing by ρ_0, and transforming the Coriolis term to the local coordinate system (see Fig. 2.31) results in the **Boussinesq equations of the ocean**:

$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u}) + \frac{\nabla p}{\rho_0} + \frac{\rho}{\rho_0} g\, \boldsymbol{k} + 2\,\Omega \begin{pmatrix} 0 & -\sin(\varphi) & \cos(\varphi) \\ \sin(\varphi) & 0 & 0 \\ -\cos(\varphi) & 0 & 0 \end{pmatrix} \boldsymbol{u} - \frac{\nabla \cdot \boldsymbol{\tau}}{\rho_0} = 0, \quad (2.17)$$

$$\nabla \cdot \boldsymbol{u} = 0. \quad (2.18)$$

Transport equations for temperature and salinity (2.15) and the equation of state (2.16) complement system (2.17)–(2.18).

- The conservation of mass represented by (2.11) is now replaced by the continuity equation (2.18) that expresses conservation of volume
- (2.18) implies a divergence-free fluid (sea water is, in fact, very slightly compressible at ca. 5×10^{-7} kg m^{-3} Pa^{-1})
- Similarly to the incompressible Navier-Stokes equations, the Boussinesq system poses a saddle-point problem

Ocean: Treatment of viscous terms

For incompressible Newtonian fluids, the viscous stress tensor

$$\boldsymbol{\tau} = -\frac{2}{3}\mu(\nabla \cdot \boldsymbol{u})\boldsymbol{I} + \mu\left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right)$$

reduces to

$$\boldsymbol{\tau} = \mu\left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right) \quad \text{or, equivalently,} \quad \frac{\boldsymbol{\tau}}{\rho_0} = \nu\left(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T\right), \quad (2.19)$$

where μ, ν denote the **dynamic and kinematic viscosity** coefficients, respectively.

Assuming a constant viscosity and using continuity equation (2.18) we have $\nabla \cdot (\nabla \boldsymbol{u})^T = 0$, and the viscous term simplifies to

$$\frac{\nabla \cdot \boldsymbol{\tau}}{\rho_0} = \nabla \cdot (\nu\, \nabla \boldsymbol{u}) = \nu \nabla^2 \boldsymbol{u} = \begin{pmatrix} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \\ \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \\ \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \end{pmatrix}, \quad (2.20)$$

where ∇^2 is called the (vector) Laplace operator.

Ocean: Dimensional analysis of the Boussinesq system
To identify geophysically relevant motion scales, we **non-dimensionalize** the Boussinesq system by introducing typical values for the horizontal and the vertical lengths and velocities denoted by L, H, U, and W, respectively. Introducing $\varepsilon := H/L$, the typical **aspect ratio** of vertical to horizontal dimensions, we note that ε rarely exceeds $1/100$ for the majority of realistic ocean domains.
Marking by prime the **dimensionless** unknowns, we can write

$$x = Lx', \qquad y = Ly', \qquad z = Hz' = \varepsilon L z', \qquad t = \frac{L}{U}t',$$
$$\nabla = \frac{1}{L}\nabla', \qquad \frac{\partial}{\partial z} = \frac{1}{\varepsilon L}\frac{\partial}{\partial z'}, \qquad \frac{\partial}{\partial t} = \frac{U}{L}\frac{\partial}{\partial t'}$$

and, analogously, for the main physical quantities

$$u = Uu', \quad v = Uv', \quad w = Ww', \quad \rho = \rho_0\rho', \quad p = \rho_0 U^2 p', \quad g = \frac{U^2}{\varepsilon L}g'.$$

Denoting by subscript $_2$ the horizontal (xy-) components of 3D vectors and operators, the continuity equation (2.18) becomes

$$\frac{U}{L}\nabla_2' \cdot \boldsymbol{u}_2' + \frac{W}{\varepsilon L}\frac{\partial w'}{\partial z'} = 0. \tag{2.21}$$

Our next goal is to identify the relationship between typical scales of vertical and horizontal velocities.

- Suppose $W \gg \varepsilon U$
- It would follow that $\frac{\partial w'}{\partial z'} \approx 0$ implying w nearly constant over the vertical extent of the domain
- No-normal-flow boundary conditions at the bottom imply $w = 0$
- Then we have $\nabla_2 \cdot \boldsymbol{u}_2 \approx 0$, and the lateral (land) boundaries (also using no-normal-flow boundary conditions) result in $\boldsymbol{u} \approx 0$ – a rather unlikely situation in a real ocean
- From now on, we assume $W = \varepsilon U$ and note that all arguments also hold for $W \ll \varepsilon U$

We introduce the dimensionless **Rossby** (Ro) and **Reynolds** (Re) numbers that play a key role in determining the character of ocean circulation

$$\mathrm{Ro} := \frac{U}{2\Omega L}, \qquad \mathrm{Re} := \frac{UL}{\nu}.$$

For the momentum equations (2.17), we separate the horizontal from the vertical momentum exchange arriving after the non-dimensionalization at

$$\frac{U^2}{L}\left\{\frac{\partial \boldsymbol{u}_2'}{\partial t'} + \nabla_2' \cdot (\boldsymbol{u}_2' \otimes \boldsymbol{u}_2') + \frac{\partial\left(\boldsymbol{u}_2' w'\right)}{\partial z'} + \nabla_2' p' + \frac{1}{\mathrm{Ro}}\begin{pmatrix} 0 & -\sin(\varphi) \\ \sin(\varphi) & 0 \end{pmatrix}\boldsymbol{u}_2' \right. \tag{2.22}$$
$$\left. + \frac{\varepsilon}{\mathrm{Ro}}\begin{pmatrix} \cos(\varphi) & 0 \\ 0 & 0 \end{pmatrix} w' - \nabla_2' \cdot \left(\frac{1}{\mathrm{Re}}\nabla_2' \boldsymbol{u}_2'\right) - \frac{\partial}{\partial z'}\left(\frac{1}{\varepsilon^2\,\mathrm{Re}}\frac{\partial \boldsymbol{u}_2'}{\partial z'}\right)\right\} = 0,$$

$$\frac{\varepsilon U^2}{L}\left\{\frac{\partial w'}{\partial t'} + \nabla_2' \cdot (\boldsymbol{u}_2' w') + \frac{\partial\left(w'^2\right)}{\partial z'} + \frac{1}{\varepsilon^2}\frac{\partial p'}{\partial z'} + \frac{1}{\varepsilon^2}\rho'\, g' \right. \tag{2.23}$$
$$\left. - \frac{1}{\varepsilon\,\mathrm{Ro}}\cos(\varphi)\, u' - \nabla_2' \cdot \left(\frac{1}{\mathrm{Re}}\nabla_2' w'\right) - \frac{\partial}{\partial z'}\left(\frac{1}{\varepsilon^2\,\mathrm{Re}}\frac{\partial w'}{\partial z'}\right)\right\} = 0.$$

Ocean: The Reynolds number and the role of viscosity

- The Reynolds number $\mathrm{Re}= \frac{UL}{\nu}$ describes the relative importance of inertial (advective) forces compared to viscous forces
- It is very large for flows on geophysical scales even when turbulence effects are taken into account
- A clear contrast between the horizontal and vertical viscosities is apparent whose coefficients may differ by several orders of magnitude
- The much larger vertical viscosity term i.e., the last term on the left-hand side of (2.22)

  $$\frac{\partial}{\partial z'}\left(\frac{1}{\varepsilon^2\,\mathrm{Re}}\frac{\partial \boldsymbol{u}_2'}{\partial z'}\right) \qquad \text{and of (2.23)} \qquad \frac{\partial}{\partial z'}\left(\frac{1}{\varepsilon^2\,\mathrm{Re}}\frac{\partial w'}{\partial z'}\right)$$

  plays an important role as a major mechanism of vertical momentum transport toward its sinks via bottom friction at the sea bed or via interaction with the wind at the free surface

Ocean: Hydrostatic balance and the vertical Coriolis force

- Starting from vertical momentum equation (2.23), omitting primes

$$\frac{\varepsilon U^2}{L}\Bigg\{\frac{\partial w}{\partial t}+\nabla_2\cdot(\boldsymbol{u}_2 w)+\frac{\partial\left(w^2\right)}{\partial z}+\frac{1}{\varepsilon^2}\frac{\partial p}{\partial z}+\frac{1}{\varepsilon^2}\rho\, g \\ -\frac{1}{\varepsilon\,\mathrm{Ro}}\cos(\varphi)\,u-\nabla_2\cdot\left(\frac{1}{\mathrm{Re}}\nabla_2 w\right)-\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon^2\,\mathrm{Re}}\frac{\partial w}{\partial z}\right)\Bigg\}=0,$$

  and collecting leading-order terms ($\sim 1/\varepsilon^{-2}$) except for the vertical viscosity term multiplying $\frac{1}{\mathrm{Re}}\ll 1$ leads to

$$\frac{\partial p}{\partial z}+\rho\, g=0, \tag{2.24}$$

  which is also known as the **hydrostatic balance** condition

- The vertical Coriolis terms (those containing $\cos(\varphi)$) tend to be much smaller than their horizontal counterparts for $\varepsilon\ll 1$ and are neglected in the following (as a part of so-called **traditional approximation**); these terms, however, can be important in non-hydrostatic settings

Ocean: The Rossby number and the geostrophic balance

- The Rossby number $\mathrm{Ro}=\frac{U}{2\Omega L}$ in (2.22) relates advection to the Coriolis force and is of order of unity or less. Horizontal Coriolis terms are very important for the slow-varying large-scale circulation in global ocean where their magnitude may exceed by far that of the advective terms

- For such types of flow, the velocity variations are usually very small, thus the horizontal pressure gradient is the only term comparable in size to the Coriolis force. Then the leading-order approximation to the horizontal momentum equations is given by

$$\nabla_2 p+\left(\begin{smallmatrix}0 & -f_c\\ f_c & 0\end{smallmatrix}\right)\boldsymbol{u}_2=0,\quad \text{where}\quad f_c\coloneqq 2\,\Omega\sin(\varphi) \tag{2.25}$$

- The above expression is called the **geostrophic balance** condition; it constitutes, similarly to the hydrostatic balance, a fundamental characteristic of large-scale ocean circulation

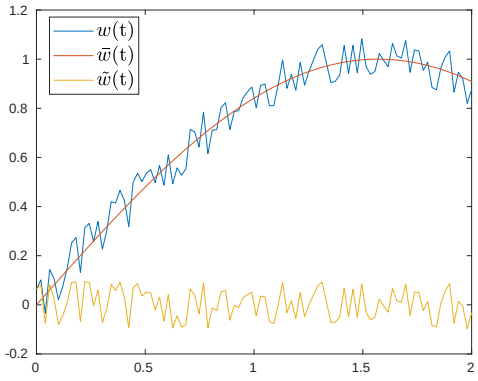
- In coastal ocean with its fast-moving flows, Coriolis terms are not as important, and pressure gradients are often balanced by the advection

Questions

- What is the main idea behind the Boussinesq approximation?
- Which conservation unknowns are different in the Boussinesq equations from the Navier-Stokes ones?
- What is a Newtonian fluid?
- How does one perform a dimensional analysis and why is it useful?
- What is the aspect ratio and what are typical aspect ratios for ocean applications?
- What relationship is expressed by the Reynolds number and why is it useful?
- What is the hydrostatic balance and why does it have a dramatic effect on the model equations?
- What is the geostrophic balance and what types of of ocean circulation are represented well by it?

Ocean: Modeling subgrid scale dynamics

- The viscous stress term $\nu\left(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T\right)$ in (2.19) is extremely small for the sea water ($\nu \approx 10^{-6}$ m^2 s^{-1}); however, terms formally similar to the molecular viscosity but much larger in size are frequently employed to model the effects of turbulent subscale motions
- The standard approach to derive these terms is based on the **Reynolds averaging** technique in which a time-varying function $\omega(t)$ is decomposed as a sum of its mean value $\overline{\omega}(t)$ assumed to be constant the averaging time period Δt and an instantaneous deviation from the mean $\widetilde{\omega}(t)$



$$\omega(t) = \overline{\omega}(t) + \widetilde{\omega}(t) \qquad (2.26)$$

The averaging $\bar{\cdot}$ operator has the following properties:

- $\overline{\omega_1 + \omega_2} = \overline{\omega}_1 + \overline{\omega}_2$,
- $\overline{\overline{\omega}} = \overline{\omega}$,
- $\overline{\widetilde{\omega}} = 0$,
- $\overline{\overline{\omega}\,\overline{\omega}} = \overline{\omega}\,\overline{\omega}$,
- $\overline{\overline{\omega}\,\widetilde{\omega}} = 0$

Splitting $\boldsymbol{u}$, p, and ρ into the mean and the deviation, substituting into momentum equation (2.17), recalling that for a constant viscosity coefficient $\nabla \cdot (\nu \nabla \boldsymbol{u}^T) = 0$, and applying the Reynolds averaging results in

$$\frac{\partial \overline{\boldsymbol{u}}}{\partial t} + \nabla \cdot (\overline{\boldsymbol{u}} \otimes \overline{\boldsymbol{u}}) + \frac{1}{\rho_0} \nabla \overline{p} + \frac{\overline{\rho}}{\rho_0} g\, \boldsymbol{k} + f_c\, \boldsymbol{k} \times \overline{\boldsymbol{u}} - \nabla \cdot \left(\nu\, \nabla \overline{\boldsymbol{u}} - \overline{\widetilde{\boldsymbol{u}} \otimes \widetilde{\boldsymbol{u}}} \right) = 0. \quad (2.27)$$

Note that all linear terms vanish due to the linearity of the averaging operator.
Equation (2.27) serves as the starting point for incorporating the **eddy viscosity** effects into the model. It has several important characteristics:

- Having two velocities ($\overline{\boldsymbol{u}}$ and $\widetilde{\boldsymbol{u}}$) instead of one without adding any new equations, poses the so-called closure problem how to express terms containing $\widetilde{\boldsymbol{u}}$ through $\overline{\boldsymbol{u}}$
- Products of deviation terms arising from the advection operator, namely $\overline{\widetilde{\boldsymbol{u}} \otimes \widetilde{\boldsymbol{u}}}$, can be interpreted as corrections to the viscous stress; this motivates the simplest possible eddy viscosity parametrization: A linear eddy viscosity coefficient of the same type as the kinematic viscosity coefficient ν but several orders of magnitude larger

Remark 2.6. *Reynolds averaging can be applied directly to the Navier–Stokes equations, the system processed in this way is called Reynolds averaged Navier–Stokes equations (RANS). Next to the Large Eddy Simulation (LES) it represents the most common way to account for turbulent effects.*

An anisotropy between the vertical and the horizontal viscosity clearly follows from the non-dimensionalization given in (2.22)–(2.23); therefore, we

emphasize this distinction by introducing separate horizontal and vertical eddy viscosity coefficients in (2.28)

$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u}) + \frac{\nabla p}{\rho_0} + \frac{\rho}{\rho_0} g\, \boldsymbol{k} + f_c\, \boldsymbol{k} \times \boldsymbol{u} - \nabla \cdot (\boldsymbol{\mathcal{D}}\, \nabla \boldsymbol{u}) = 0, \qquad (2.28)$$

where $\boldsymbol{\mathcal{D}}$ is the tensor of eddy viscosity coefficients defined as

$$\boldsymbol{\mathcal{D}} = \begin{pmatrix} D_u & 0 & 0 \\ 0 & D_v & 0 \\ 0 & 0 & D_w \end{pmatrix}, \quad D_u = D_v = D_w = \begin{pmatrix} A_u & 0 & 0 \\ 0 & A_u & 0 \\ 0 & 0 & \nu_t \end{pmatrix}, \quad \boldsymbol{\mathcal{D}}\, \nabla \boldsymbol{u} := \begin{pmatrix} (D_u \nabla u)^T \\ (D_v \nabla v)^T \\ (D_w \nabla w)^T \end{pmatrix}$$

where A_u is the horizontal and ν_t the vertical eddy viscosity coefficient.

Remark 2.7. (2.28) *is, in fact, an equation for* $\overline{\boldsymbol{u}}$ *although we drop the bar for brevity.*

Remark 2.8. *The horizontal and the vertical eddy viscosity coefficients express physically distinct momentum exchange mechanisms and can differ by several orders of magnitude.*

In a similar way, one obtains transport equations for temperature and salinity with **eddy diffusivity** terms:

$$\frac{\partial r}{\partial t} + \nabla \cdot (\boldsymbol{u}\, r) - \nabla \cdot (D_r\, \nabla r) = 0, \quad D_r = \begin{pmatrix} A_r & 0 & 0 \\ 0 & A_r & 0 \\ 0 & 0 & \nu_r \end{pmatrix}, \quad r \in \{T, S\}\,. \qquad (2.29)$$

Questions

- Describe the main idea of Reynolds averaging.
- What are the main properties of the averaging operator?
- How are the eddy viscosity (turbulence) effects incorporated in the momentum equations?
- Why is there a clear difference between the eddy viscosity mechanisms in the horizontal and vertical directions?

Ocean: The primitive (hydrostatic) equations
The non-dimensionalzied Boussinesq vertical momentum equation (2.23)

$$\frac{\partial w'}{\partial t'} + \nabla' \cdot (\boldsymbol{u}' w') + \frac{1}{\varepsilon^2}\left(\frac{\partial p'}{\partial z'} + \rho'\, g'\right) - \frac{1}{\varepsilon\,\mathrm{Ro}} \cos(\varphi)\, u' - \nabla_2' \cdot \left(\frac{1}{\mathrm{Re}} \nabla_2' w'\right) - \frac{\partial}{\partial z'}\left(\frac{1}{\varepsilon^2\,\mathrm{Re}} \frac{\partial w'}{\partial z'}\right) = 0$$

leads to

Idea of the hydrostatic approximation

According to the dimensional analysis, for $\varepsilon \ll 1$, the retention of only leading-order terms reduces the vertical momentum equation to the statement of hydrostatic balance given in (2.24)

$$\frac{\partial p}{\partial z} + \rho\, g = 0.$$

Taking the above equality as a **diagnostic** equation for determining the pressure dramatically changes the mathematical character of the system and has profound implications from the physical point of view.

The **hydrostatic equations of the ocean** is the most commonly used mathematical model for simulating the **baroclinic** (variable density) circulation in global, regional, and coastal ocean. They are given by

$$\frac{\partial \boldsymbol{u}_2}{\partial t} + \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u}_2) + \frac{\nabla_2 p}{\rho_0} + \begin{pmatrix} 0 & -f_c \\ f_c & 0 \end{pmatrix} \boldsymbol{u}_2 - \nabla \cdot (\boldsymbol{\mathcal{D}}_2\, \nabla \boldsymbol{u}_2) = 0, \tag{2.30}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{2.31}$$

$$\frac{\partial p}{\partial z} + \rho\, g = 0, \tag{2.32}$$

$$\rho = \rho(T, S, p), \tag{2.33}$$

$$\frac{\partial r}{\partial t} + \nabla \cdot (\boldsymbol{u} r) - \nabla \cdot (D_r\, \nabla r) = 0, \quad r \in \{T, S\}\,. \tag{2.34}$$

System (2.30)–(2.34) is complemented by the **kinematic** and **dynamic** boundary conditions.
We only state the conditions at the free surface and the sea bed for the continuity and momentum equations; the boundary conditions at the lateral boundaries come in more varieties and can be found in many textbooks on ocean modeling.
Denoting by $\boldsymbol{n} = (n_x, n_y, n_z)^T$ an exterior unit normal to the ocean domain, we specify at the **bottom (sea bed) boundary**

- Kinematic: No-normal-flow

$$\boldsymbol{u}_{|_{z=z_b}} \cdot \boldsymbol{n} = 0 \tag{2.35}$$

- Dynamic: Quadratic slip for the horizontal velocity components

$$\boldsymbol{\mathcal{D}}_2 \nabla \boldsymbol{u}_{2|_{z=z_b}} \cdot \boldsymbol{n} = -C_f \left|\boldsymbol{u}_{|_{z=z_b}}\right| \boldsymbol{u}_{2|_{z=z_b}}, \tag{2.36}$$

  where z_b denotes the value of the z-coordinate at the sea bed, and $C_f > 0$ is a constant friction coefficient

The standard boundary conditions at the **free surface** have the form (here, ξ denotes the value of the z-coordinate at the free surface):

- Kinematic (points at the free surface move with it)

$$\frac{\partial \xi}{\partial t} + u_{|_{z=\xi}} \frac{\partial \xi}{\partial x} + v_{|_{z=\xi}} \frac{\partial \xi}{\partial y} - w_{|_{z=\xi}} = 0 \tag{2.37}$$

- Dynamic: Pressure and wind stress are prescribed

$$p_{|_{z=\xi}} = p_{\mathrm{a}}, \tag{2.38}$$

  where p_{a} is the atmospheric pressure, and

$$\boldsymbol{\mathcal{D}}_2 \nabla \boldsymbol{u}_{2|_{z=\xi}} \cdot \boldsymbol{n} = \frac{\boldsymbol{\tau}_s}{\rho_0}, \tag{2.39}$$

  with $\boldsymbol{\tau}_s$ denoting the surface wind stress

Their main properties can be summarized as follows:

- Pressure is computed from the **hydrostatic pressure condition** $\frac{\partial p}{\partial z} + \rho\, g = 0$ and only accounts for the hydrostatic (weight of the water above the point) and not for the hydrodynamic effects

- The **prognostic** vertical momentum equation for the vertical velocity w is not included in the hydrostatic system, and w is computed diagnostically using the continuity equation (2.31) $\nabla \cdot \boldsymbol{u} = 0$

- As opposed to the non-hydrostatic Boussinesq system, **the hydrostatic equations do not pose a saddle point problem**

- Using equation (2.32) and the pressure boundary condition at the free surface (2.38), the hydrostatic pressure gradient in the momentum equation (2.30) is usually represented in the following convenient form

$$\frac{1}{\rho_0}\nabla_2 p = \frac{1}{\rho_0}\nabla_2\left(p_a + \int_z^\xi \rho\, g\, d\zeta\right) = \frac{\nabla_2 p_a}{\rho_0} + g\,\nabla_2\,\xi + \frac{g}{\rho_0}\nabla_2\int_z^\xi (\rho - \rho_0)\, d\zeta$$

- While much better treatable numerically, from the physical point of view, the hydrostatic system presents a number of **limitations**:
  - Vertical accelerations are neglected
  - The vertical momentum is not conserved
  - Vertical accelerations affect the propagation of internal and short surface waves, therefore any studies sensitive to an accurate representation of these phenomena require a non-hydrostatic model
- In summary, any problem setting focusing on domains and flow regimes with the ratio of vertical-to-horizontal scales greater than 1/10 most likely should be modeled using a non-hydrostatic system

Ocean: 2D shallow water equations

- Even in the substantially simplified settings of non-hydrostatic (Boussinesq) or hydrostatic (primitive) equations, the resulting PDE system is still rather complex and computationally expensive to solve
- For problems not greatly affected by the density-driven dynamics but focusing instead on fast moving surface waves (tidal flows, tsunamis), one can make further simplifications to obtain a 2D system of PDEs
- By integrating the continuity equation $\nabla \cdot \boldsymbol{u} = 0$ vertically over the depth and applying kinematic boundary conditions at the free surface

$$\frac{\partial \xi}{\partial t} + u_{|_{z=\xi}}\frac{\partial \xi}{\partial x} + v_{|_{z=\xi}}\frac{\partial \xi}{\partial y} - w_{|_{z=\xi}} = 0$$

and the sea bed $\boldsymbol{u}_{|_{z=z_b}} \cdot \boldsymbol{n} = 0$,[1ex] we obtain the 2D (primitive) continuity equation

$$\frac{\partial \xi}{\partial t} + \nabla_2 \cdot \int_{z_b}^{\xi} \boldsymbol{u}_2\, dz = 0 \tag{2.40}$$

The treatment of the momentum equation (2.30)

$$\frac{\partial \boldsymbol{u}_2}{\partial t} + \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u}_2) + \frac{\nabla_2 p}{\rho_0} + \left(\begin{smallmatrix} 0 & -f_c \\ f_c & 0 \end{smallmatrix}\right) \boldsymbol{u}_2 - \nabla \cdot (\boldsymbol{\mathcal{D}}_2 \nabla \boldsymbol{u}_2) = 0$$

is somewhat more involved. We proceed as follows:

- Assume constant density $\rho = \rho_0$ and a vertically uniform horizontal velocity $\boldsymbol{u}_2$[5]
- Denote by $\boldsymbol{q} \coloneqq \int_{z_b}^{\xi} \boldsymbol{u}_2 \, dz$ the depth integrated horizontal velocity and by $H \coloneqq \xi - z_b$ the total water depth
- Next, the momentum equation (2.30) is integrated vertically applying the dynamic boundary conditions at the free surface

  $$\boldsymbol{\mathcal{D}}_2 \nabla \boldsymbol{u}_{2|_{z=\xi}} \cdot \boldsymbol{n} = \boldsymbol{\tau}_s / \rho_0$$

  and at the sea bed

  $$\boldsymbol{\mathcal{D}}_2 \nabla \boldsymbol{u}_{2|_{z=z_b}} \cdot \boldsymbol{n} = -C_f \, |\boldsymbol{u}_{|_{z=z_b}}| \, \boldsymbol{u}_{2|_{z=z_b}}$$

Skipping some intricacies of the viscosity parametrization, we end up with the system of two-dimensional (viscous) shallow water equations

$$\frac{\partial \boldsymbol{q}}{\partial t} + \nabla_2 \cdot \frac{\boldsymbol{q} \otimes \boldsymbol{q}}{H} + gH \nabla_2 \, \xi + \left(\begin{smallmatrix} 0 & -f_c \\ f_c & 0 \end{smallmatrix}\right) \boldsymbol{q} - \frac{C_f^*}{H} |\boldsymbol{q}| \, \boldsymbol{q} - \frac{\boldsymbol{\tau}_s}{\rho_0} - \nabla_2 \cdot (A_u \nabla_2 \, \boldsymbol{q}) = 0, \tag{2.41}$$

$$\frac{\partial \xi}{\partial t} + \nabla_2 \cdot \boldsymbol{q} = 0, \tag{2.42}$$

where $C_f^* = \text{const} > 0$ denotes the bottom friction coefficient for the depth-integrated momentum equations.

Remark 2.9. *Equation* (2.42) *is often utilized in 3D hydrostatic and non-hydrostatic systems in place of* (2.37) *to determine the free surface elevation. The usual way to accomplish this is called mode-splitting and boils down to solving system* (2.41), (2.42) *to obtain the free surface elevation that is then used in the solution of the full 3D system to determine the 3D fields.*

[5]These assumptions are reasonably realistic for shallow coastal shelves dominated by tidal dynamics where fast flows result in an efficient vertical mixing.

2.7 Atmosphere

2.7.1 Physics of the atmosphere

Atmosphere: Chemical composition

- **Dry air** is mainly composed of nitrogen (78.08 % in volume), oxygen (20.95% in volume), argon (0.93% in volume) and to a lesser extent carbon dioxide (380 ppm or 0.038% in volume)
- The remaining fraction is made up of various trace constituents such as neon (18 ppm), helium (5 ppm), methane (1.75 ppm), and krypton (1 ppm)
- In addition, a highly variable amount of water vapor is present in the air. This ranges from approximately 0% in the coldest part of the atmosphere to as much as 5% in moist and hot regions. On average, water vapor accounts for 0.25% of the mass of the atmosphere

Atmosphere: Equation of state
The equation of state is based on the ideal gas law

$$p = \rho R T,$$

where p is the pressure, ρ the density, R the universal gas constant ($R \approx 8.31287$[J/(K mol)]), and T the absolute temperature.
Similar relationships also hold for **dry air** (subscript $_d$) and **water vapor** (subscript $_v$)

$$p_d = \rho_d R_d T, \qquad p_v = \rho_v R_v T,$$

where R_d and R_v denote the specific gas constants obtained using the molecular masses of dry air m_d and water vapor m_v via

$$R_d = R/m_d, \qquad R_v = R/m_v.$$

The total air pressure and density are then given as sums of those of dry air and water vapor

$$p = p_d + p_v, \qquad \rho = \rho_d + \rho_v.$$

Introducing the specific humidity $q = \rho_v/\rho$ and the ratio between the molecular masses of dry air and water vapor $\varepsilon = m_v/m_d$, the equation of state for moist air can be expressed as

$$\rho(p, T, q) = \frac{p}{R_d T(1 - q + \frac{q}{\varepsilon})} = \frac{p}{R_m T}, \tag{2.43}$$

where R_m is the specific gas constant for moist air

$$R_m = R_d \left(1 - q + \frac{q}{\varepsilon}\right) = \frac{R}{m_d} \left(1 - q + \frac{q}{\varepsilon}\right).$$

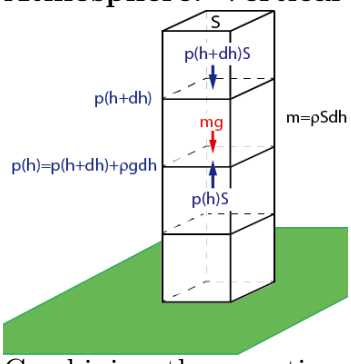
Denoting by

$$\alpha = \frac{1}{\rho}$$

the **specific volume** of gas, (2.43) can be equivalently written as

$$p\,\alpha = R_m T. \tag{2.44}$$

Atmosphere: Vertical pressure profile



On a large-scale, the atmosphere is nearly in hydrostatic equilibrium, meaning that at height z, the force due to the pressure p on a 1 m^2 horizontal surface balances the force due to the weight of the air above z. The atmospheric pressure is thus at its maximum at the Earth's surface, and the surface pressure p_s depends on the mass of the air column at a particular location

Combining the equation of state for moist air (2.43) with the hydrostatic pressure $\partial p/\partial z = -\rho g$ results in an ODE for pressure

$$\frac{\partial p}{\partial z} = -\frac{p}{R_m T} g, \; p|_{z=0} = p_s \quad \Rightarrow \quad p = p_s e^{-\frac{g\,z}{R_m T}} = p_s e^{-\frac{z}{H}},$$

where $H = R_m T/g$ is a **scale height**[6] (ca. 8 km for the atmosphere).

[6]Distance over which a quantity (pressure in our case) decreases by a factor of e.

Atmosphere: The first law of thermodynamics
Application of the first law of thermodynamics to a parcel of fluid relates the heat (energy) exchange with the surroundings dQ, the internal energy change dU, and the work done by or on the parcel dW

$$dQ = dU + dW,$$

where d denotes an infinitesimal change in the corresponding quantity.

- For a parcel of gas of mass M, if the volume V changes then there is some work done

  $$dW = p\frac{dV}{M} = p\,d\alpha,$$

  where $\alpha = 1/\rho$ denotes once again the specific volume.

- The change in internal energy is proportional to the temperature change for a fixed volume

  $$dU = C_v dT,$$

  where C_v is the specific heat of moist air at constant volume

Atmosphere: Adiabatic processes
Combining the expressions for internal energy and work, we arrive at

$$dQ = C_v dT + p\,d\alpha. \tag{2.45}$$

Differentiating the equation of state in the form (2.44) $p\alpha = R_m T$ gives

$$p\,d\alpha + \alpha dp = R_m dT$$

which can be then substituted into (2.45) resulting in

$$dQ = C_p dT - \alpha dp, \tag{2.46}$$

where $C_p = C_v + R_m$ is the specific heat of moist air at constant pressure.
A thermodynamic process is called **adiabatic** if it takes place without exchanging heat with the environment, i.a.

$$dQ = 0 \qquad \Leftrightarrow \qquad C_p dT - \alpha dp = 0. \tag{2.47}$$

Since the thermal conductivity of air is very small, many processes that do not involve radiative transfer can be considered adiabatic.

Atmosphere: Lapse rate
Substituting the hydrostatic pressure $\partial p/\partial z = -\rho g$ into (2.47) gives

$$C_p dT = -\alpha\,\rho\,g\,dz \qquad \Leftrightarrow \qquad -\frac{\partial T}{\partial z} = \frac{g}{C_p}$$

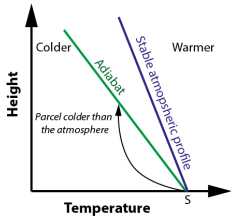
motivating the definition of $\Gamma_d := \frac{g}{C_p}$ called the adiabatic **lapse rate** and describing the rate of change of temperature with altitude (for adiabatic and unsaturated conditions).

- The temperature in the **troposphere**, roughly the lowest 10 km of the atmosphere, generally decreases with height
- The measured lapse rate $\Gamma := -\frac{\partial T}{\partial z}$ depends on the radiative balance of the atmosphere, convection, and the horizontal heat transport. Its global mean value is around 6.5 K km^{-1}, but Γ depends on the location and season
- The lapse rate is an important characteristic of the atmosphere. For instance, it determines its **static stability**

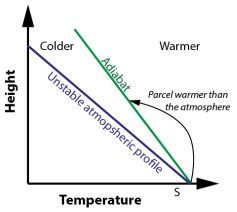
Atmosphere: Stable and unstable conditions

Consider a parcel of air rising adiabatically. Its temperature changes according to the adiabatic lapse rate Γ_d, and its pressure is equal to the air pressure at the current altitude. However, the temperature (and thus also the density) of the surrounding air changes according to the measured lapse rate Γ, and those can be different from that of the rising air parcel.

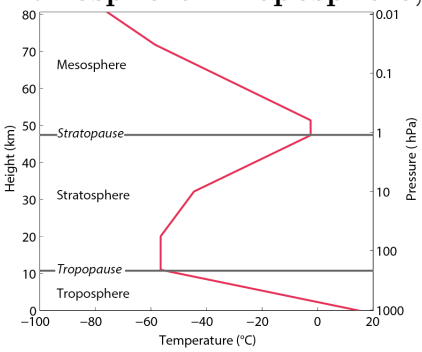
Low measured lapse rates inhibit vertical movements and increase the vertical stability. Negative lapse rates (temperature increasing with height), called temperature inversions, correspond to highly stable conditions



When the measured lapse rate rises, a vertical instability followed by convection may arise. The lapse rate is also involved in feedbacks and plays an important role in the response of the atmosphere to perturbations



Atmosphere: Troposphere, Stratosphere, Mesosphere



At an altitude of about 10 km, a region of weak vertical temperature gradients, called the **tropopause**, separates the troposphere from the stratosphere where the temperature generally increases with height until the **stratopause** at around 50 km.
In the **mesosphere**, temperature decreases strongly with height until ca. 80 km and then increases again in the **thermosphere**.

The vertical gradients above 10 km are strongly influenced by the absorption of solar radiation by different atmospheric constituents and by chemical reactions driven by the incoming light. In particular, the warming in the stratosphere at heights of about 30-50 km is mostly due to the absorption of ultraviolet radiation by stratospheric ozone, which protects life on Earth from this dangerous radiation.

Atmosphere: Horizontal distribution of temperature

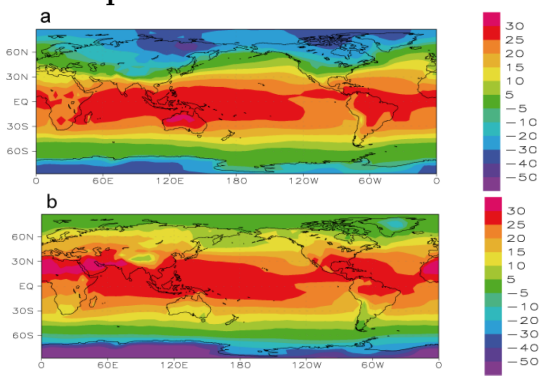


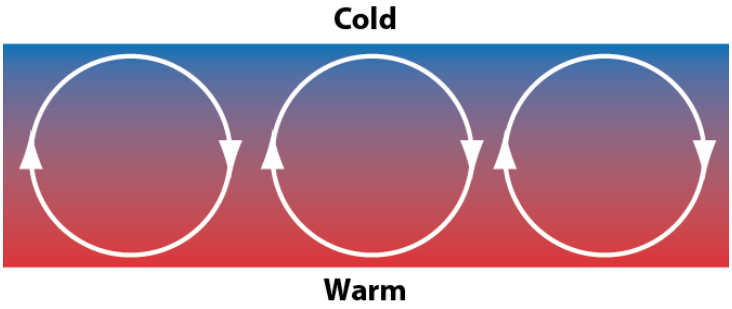
Figure 2.46: Surface air temperature [°C] (Brohan et al. 2005) averaged over
(a) December, January, and February
(b) June, July, and August

- At the Earth's surface, the temperature reaches its maximum in equatorial regions due to the higher incoming solar radiation. In those regions, the temperature is relatively constant throughout the year
- Because of the much stronger seasonal cycle at mid and high latitudes, the north-south gradients are much larger in winter than in summer

The distribution of the surface temperature is also affected by atmospheric and oceanic heat transport as well as by the thermal inertia of the ocean

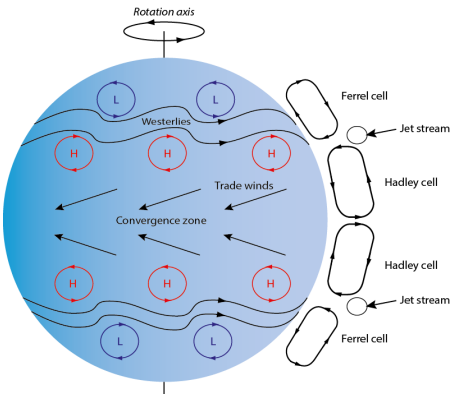
Atmosphere: Convection

- The high temperatures at the equator make the air there less dense. It thus tends to rise before being transported poleward at high altitudes in the troposphere. This motion is compensated for at the surface by an equatorward displacement of the air.
- On a motionless Earth, this big convection cell would reach the poles, inducing direct exchanges between the warmest and coldest places on Earth. Because of the Earth's rotation, such an atmospheric structure would be unstable



Atmosphere: Hadley and Ferrel cells

- Two cells driven by the ascendance at the equator, called the **Hadley cells** close with a downward branch at a latitude of about 30°C. The northern boundary of these cells is marked by strong westerly winds in the upper troposphere called the **tropospheric jets**



- The extratropical circulation is dominated at the surface by westerly winds whose zonal symmetry is perturbed by large wave-like patterns and the continuous succession of disturbances that governs the day-to-day variations in the weather in these regions. The dominant feature of the meridional circulation at those latitudes is the **Ferrell cell**, which is weaker than the Hadley cell

Atmosphere: Surface winds

- At the surface, the Earth's rotation is responsible for a deflection toward the right in the northern hemisphere and toward the left in the southern hemisphere (due to the Coriolis force) of the flow coming from the mid-latitudes to the equator. This gives rise to the easterly trade winds characteristic of the tropical regions

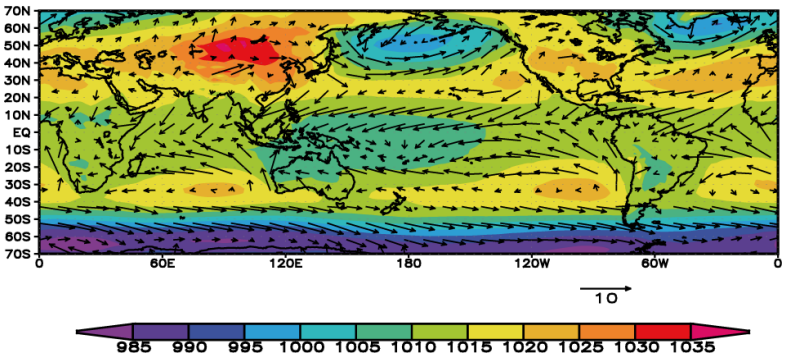


Figure 2.47: Winds [m/s] at 10m above the sea level (arrows) and sea level pressure [hPa] (colors) in December, January, and February (Kalnay et al. 1996).

Atmosphere: Precipitation

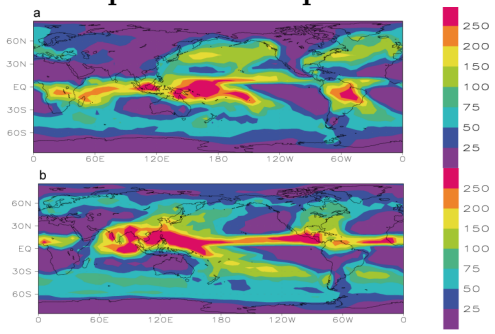


Figure 2.48: Precipitation [cm/year] (P. Xie and P. A. Arkin 1997):
(a) December, January, February;
(b) June, July, August.

The large-scale atmospheric circulation has a strong influence on precipitation, which is, with temperature, the most important variable in defining regional climate.

Along the equatorial belt, the cooling of warm and moist surface air during its rising motion leads to condensation and heavy precipitation in this area. For instance, the western tropical Pacific receives more than 3 m of rainfall per year.

By contrast, the downward motion in the subtropics is associated with the presence of very dry air and very low precipitation rates. Thus the majority of large deserts are located in the sub-tropical belt.

Atmosphere: Monsoons

- The presence of land surfaces has a critical role in **monsoon** circulation. In summer, the continents warm faster than the oceans because of their lower thermal inertia. This induces a warming of the air close to the surface and a decrease in surface pressure there inducing a transport of moist air from the sea to the land
- In winter, the situation reverses with high pressure over the cold continent and a flow generally from land to sea. Such a monsoonal circulation, with seasonal reversals of the wind direction, is present in many tropical areas of Africa, Asia and Australia. Nevertheless, the most famous monsoon is probably the South Asian one that strongly affects the Indian sub-continent
- The monsoon strongly affects precipitation in subtropics. During the winter monsoon, the inflow of dry continental air is associated with low precipitation, whereas the summer brings moist air from the ocean inducing rainfall that can reach several meters in a few months

Questions

- What kind of equation is used as the equation of state for the atmosphere?
- What is the specific volume of gas and why does one introduce it?
- What is the scale height, and under what assumption can it be calculated for the Earth's atmosphere?
- Formulate the first law of thermodynamics.
- How does one compute
  - the work of a parcel of air?
  - the change of internal energy in a parcel of air?
- What is the adiabatic lapse rate and how is it related to the measured lapse rate?
- How is the static stability of an air column related to the lapse rate?

- Name the main parts in the vertical structure of the atmosphere and describe criteria used for this classification.
- What are the convection cells and how do they arise?

2.7.2 Atmospheric modeling

Atmosphere: Energy conservation

- One of the key differences between atmospheric and oceanic GCMs is the form of the equation for energy conservation: In the latter, this is represented by a temperature transport equation, whereas the former considers a much more elaborate relationship that accounts for several important physical processes.
- The energy conservation equation for a parcel of air is called the **thermodynamic energy equation**; it is based on the first law of thermodynamics but has, in particular, to account for radiative energy fluxes and **latent heat** exchanges associated with phase changes of water (solid, liquid, gaseous).
- For the atmosphere, it can be written (cf. (2.46)) as

$$C_p \frac{DT}{Dt} - \frac{1}{\rho} \frac{Dp}{Dt} = Q, \tag{2.48}$$

  where Q denotes the heat flux (also called **diabatic heating rate**).

Atmosphere: Diabatic heating

The diabatic heating rate Q is usually represented in the following form:

$$Q = Q_{ce} + Q_{fm} + Q_{ds} + Q_{sol} + Q_{ir},$$

where

- Q_{ce} is the rate of energy release (absorption) due to condensation (evaporation)
- Q_{fm} is the rate of energy release (absorption) due to freezing (melting)
- Q_{ds} is the rate of energy release (absorption) due to deposition (sublimation)
- Q_{sol} is the rate of solar heating
- Q_{ir} is the rate of net infrared heating (cooling)

Atmosphere: Non-hydrostatic equations
The non-hydrostatic (also called **elastic**) system of equations for the atmosphere contains seven equations/unknowns

- The conservation of mass: $\dfrac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{u}) = 0$
- The conservation of mass of water vapor:
$$\frac{\partial \rho q}{\partial t} + \nabla \cdot (\rho q \boldsymbol{u}) = E - C,$$
where E and C are the evaporation and condensation, respectively
- The conservation of momentum (F represents the friction forces):
$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \cdot (\boldsymbol{u} \otimes \boldsymbol{u}) + \frac{\nabla p}{\rho} + g\boldsymbol{k} + 2\boldsymbol{\Omega} \times \boldsymbol{u} = F$$
- The conservation of energy:
$$C_p \left(\frac{\partial T}{\partial t} + \nabla \cdot (T\boldsymbol{u}) \right) - \frac{1}{\rho} \left(\frac{\partial p}{\partial t} + \nabla \cdot (p\boldsymbol{u}) \right) = Q$$
- The equation of state: $p = \rho R_m T$

Atmosphere: Other systems

- Similarly to the oceanic equations, the atmospheric ones can be simplified using the Boussinesq and hydrostatic approximations
- Another common approach considers **potential temperature**[7] θ instead of T as a primary unknown
- The relationship between temperature and potential temperature can be derived by substituting the equation of state for moist air (2.43) into the first law of thermodynamics for adiabatic conditions (2.47)
$$\frac{dT}{T} = \frac{R_m}{C_p} \frac{dp}{p} \qquad \Rightarrow \qquad T = C p^{\frac{R_m}{C_p}},$$
where C is an integration constant calculated from the definition of θ
$$\theta = T|_{p=p_s} \qquad \Rightarrow \qquad C = \theta p_s^{-\frac{R_m}{C_p}} \qquad \Rightarrow \qquad T = \theta \left(\frac{p}{p_s} \right)^{\frac{R_m}{C_p}}$$

[7] Potential temperature is the temperature an unsaturated air parcel attains if it is brought adiabatically from its altitude down to the mean sea level pressure $p_s = 10^5$ Pa

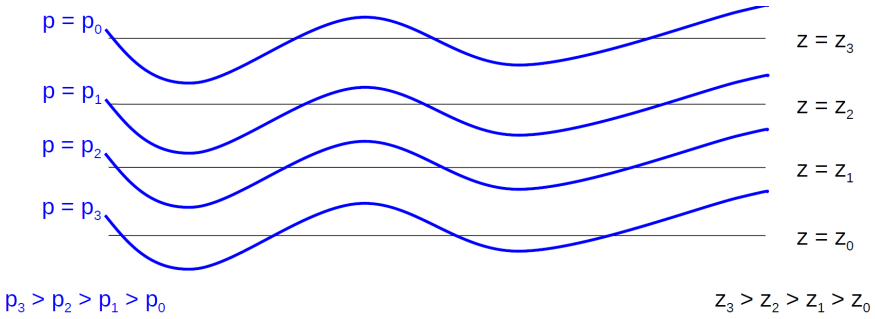


Figure 2.49: Comparison of pressure and altitude coordinates.

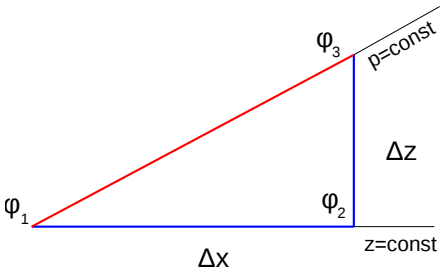
Atmosphere: Pressure coordinate

- Under the hydrostatic assumption $\partial p/\partial z = -\rho g < 0$, pressure is a monotonic function of altitude $p(z)$ and has an inverse function that expresses altitude as a monotonic function of pressure: $z(p) = p^{-1}(z)$
- This allows to formulate the hydrostatic system in the so-called pressure coordinate where the isobaric surfaces represent levels of constant vertical coordinate

Atmosphere: Derivatives in pressure coordinate

To transform spatial and temporal derivatives of a scalar field φ from standard Cartesian to pressure coordinate consider a simple relation

$$\frac{\varphi_3 - \varphi_1}{\Delta x} = \frac{\varphi_2 - \varphi_1}{\Delta x} + \frac{\varphi_3 - \varphi_2}{\Delta x} = \frac{\varphi_2 - \varphi_1}{\Delta x} + \frac{\varphi_3 - \varphi_2}{\Delta z}\frac{\Delta z}{\Delta x}.$$



Taking a limit as $\Delta x \to 0$, $\Delta z \to 0$ and denoting the derivatives on constant z-surfaces with the subscript $_z$ and on constant p-surfaces with the subscript $_p$, we obtain

$$\left(\frac{\partial \varphi}{\partial x}\right)_p = \left(\frac{\partial \varphi}{\partial x}\right)_z + \frac{\partial \varphi}{\partial z}\left(\frac{\partial z}{\partial x}\right)_p.$$

In a similar manner $$\left(\frac{\partial \varphi}{\partial y}\right)_p = \left(\frac{\partial \varphi}{\partial y}\right)_z + \frac{\partial \varphi}{\partial z}\left(\frac{\partial z}{\partial y}\right)_p.$$

Denoting the horizontal gradient operators in both coordinates by

$$\nabla_p = \begin{pmatrix} \left(\frac{\partial}{\partial x}\right)_p \\ \left(\frac{\partial}{\partial y}\right)_p \end{pmatrix}, \qquad \nabla_z = \begin{pmatrix} \left(\frac{\partial}{\partial x}\right)_z \\ \left(\frac{\partial}{\partial y}\right)_z \end{pmatrix},$$

we obtain the following transformation rules:

- $\nabla_p \varphi = \nabla_z \varphi + \dfrac{\partial \varphi}{\partial z} \nabla_p z$
- $\nabla_z \varphi = \nabla_p \varphi + \dfrac{\partial \varphi}{\partial p} \nabla_z p$
- $\left(\dfrac{\partial \varphi}{\partial t}\right)_p = \left(\dfrac{\partial \varphi}{\partial t}\right)_z + \left(\dfrac{\partial \varphi}{\partial z}\right)_t \left(\dfrac{\partial z}{\partial t}\right)_p$
- $\left(\dfrac{\partial \varphi}{\partial t}\right)_z = \left(\dfrac{\partial \varphi}{\partial t}\right)_p + \left(\dfrac{\partial \varphi}{\partial p}\right)_t \left(\dfrac{\partial p}{\partial t}\right)_z$

Atmosphere: Using pressure coordinate
Denoting the velocity in altitude coordinate as $\boldsymbol{u} = (u, v, w)$ and in pressure coordinate as $\boldsymbol{u} = (u, v, \omega)$ The vertical velocity in pressure coordinate is expressed as

$$\omega = \left(\frac{Dp}{Dt}\right).$$

This definition of vertical velocity only considers the movement relative to the movement of the material points.
The material derivative is transformed in a similar way using the hydrostatic pressure relation

$$\left(\frac{D\varphi}{Dt}\right)_z = \left(\frac{D\varphi}{Dt}\right)_p + \boldsymbol{u}_2 \cdot \nabla_p \varphi + \omega \frac{\partial \varphi}{\partial p},$$

where we use notation $\boldsymbol{u}_2 = (u, v)$.

Atmosphere: Mass conservation in pressure coordinate
Using the mass conservation equation in material derivative form

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \boldsymbol{u} = 0$$

allows to quickly transform it to pressure coordinate

$$\nabla_p \cdot \boldsymbol{u}_2 + \frac{\partial \omega}{\partial p} = \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)_p + \frac{\partial \omega}{\partial p} = 0.$$

This form of the mass conservation equation has the advantage that it does not contain density!

Questions

- What are the main differences between the energy conservation equations for the ocean and the atmosphere?
- What physical processes are usually represented by the diabatic heating rate in the energy conservation equation for the atmosphere?
- What are the main differences between the non-hydrostatic systems for the ocean and the atmosphere?
- Explain the idea of potential temperature. What are the advantages of using it?
- What is pressure coordinate? Why is it useful?

3 Environmental models

3.1 Subsurface

Modeling of flow, transport, and reactions in subsurface
With thanks to Dr. Alexander Prechtel
The modeling activities concerning the subsurface generally focus on two main processes/applications:

- Groundwater flow
- Transport and reactions of various substances carried by the groundwater

Ground water management

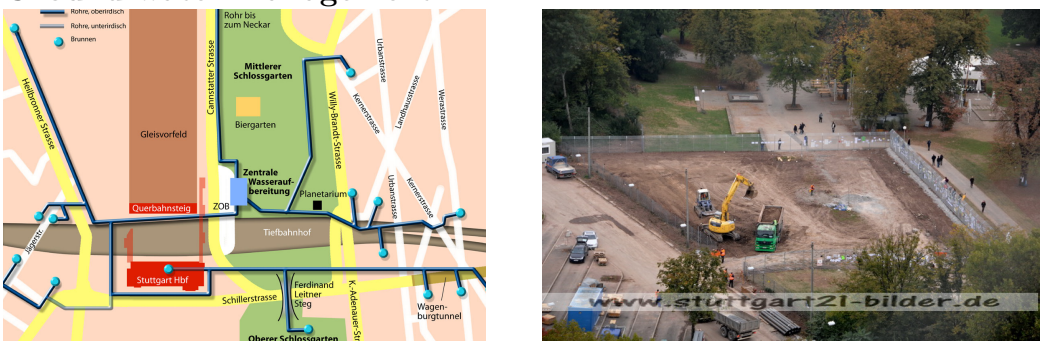
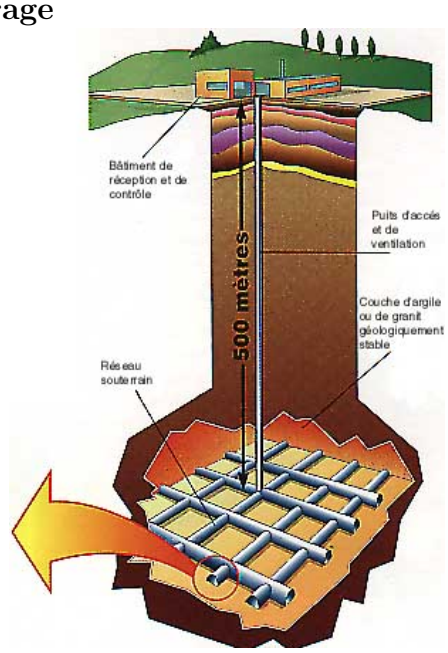


Figure 3.1: Lowering of the groundwater level from 4 to 15 meters by means of 90 wells for Stuttgart 21 construction project.

Contaminant remediation (Altlastensanierung)

Nuclear waste storage



Carbon Capture & Storage

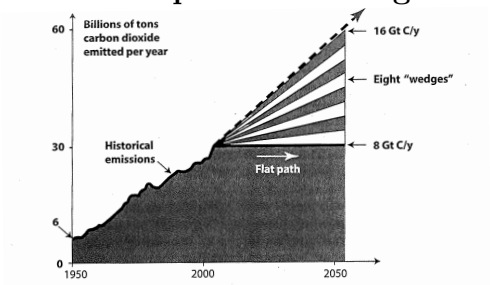


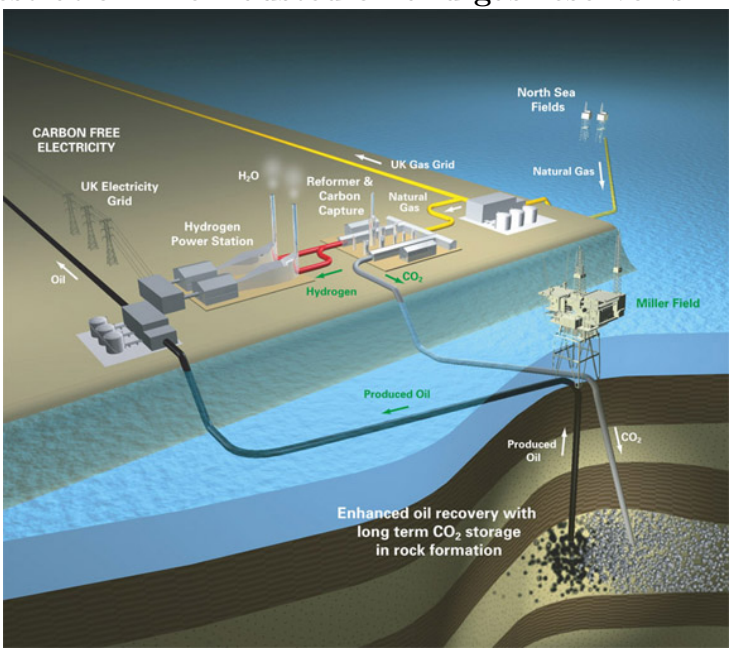
Figure 1.4 Stabilization wedges, modified from the original presented in Pacala and Socolow (2004), which used a "current" emission rate of 7 Gt C/year.

Table 1.1 Examples of Technologies and Associated Efforts to Achieve One Wedge (Updated from Pacala and Socolow 2004)

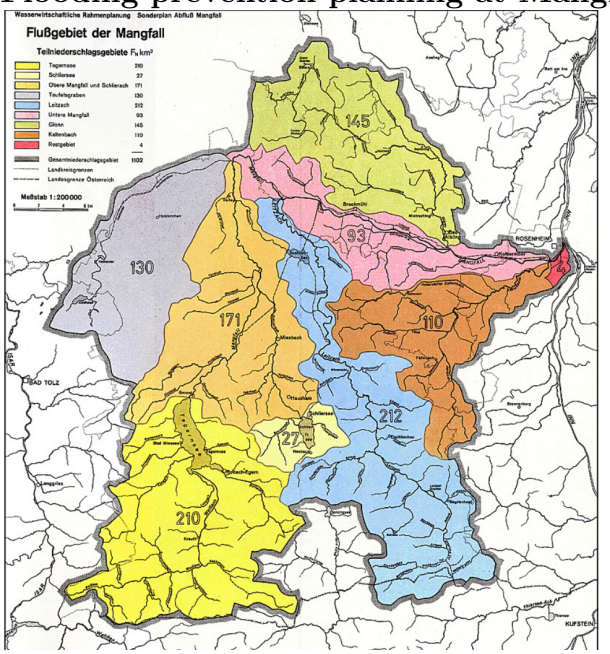
Technology	Effort required to achieve one wedge
Automobile efficiency	Increase fuel efficiency from 30 mpg to 60 mpg for 2 billion cars (currently there are fewer than 1 billion cars worldwide).
Nuclear power	Add twice the currently installed nuclear capacity.
Solar power	Add 350 times the currently installed solar power capacity.
Wind power	Add 15 times the currently installed capacity of wind-generated electricity.
Carbon capture and storage	Install carbon capture and storage at 800 large-scale coal-fired power plants.

CCS is the critical enabling technology that would reduce CO_2 emissions significantly while also allowing coal to meet the world's pressing energy needs (MIT 2007).

CO_2-Sequestration in exhausted oil and gas reservoirs



Flooding prevention planning at Mangfall



- 1100 km^2 watershed area, partially by seasonal tributaries
- mean outflow to Inn: 17,5 m^3/s
- Centennial Flood: Potential damage of ca. 1 bln. Euro, 42,000 people in the endangered area
- Prediction for the water levels
- Flooding scenarios

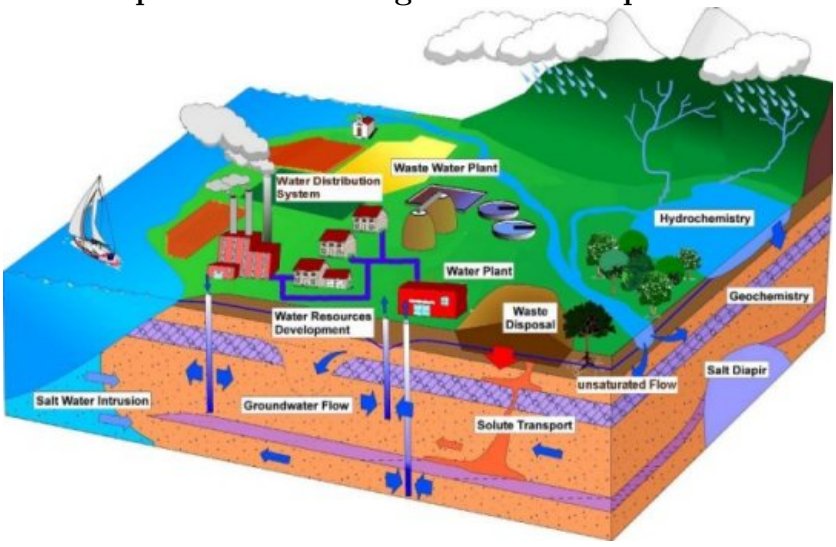
Applications of groundwater flow models

- Evaluation of pumping efforts
- Quantitative risk assessment
- Planning of remediation activities
- Prediction of groundwater level developments
- Determination of water budgets
- Physically justified approximation based on point-measurements
- Foundation for transport and reaction modeling

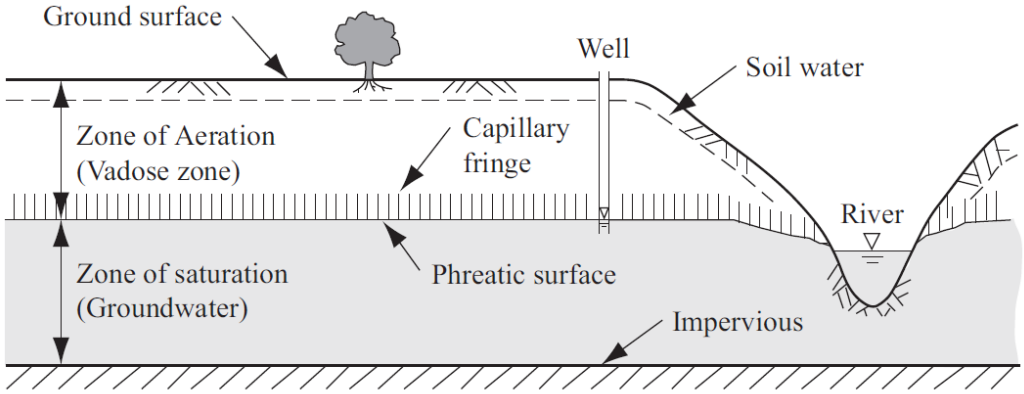
Applications of species transport models

- Interpretation of measured concentration data
- Estimation of contaminant volumes
- Planning for protection and remediation activities
- Planning and optimization of monitoring
- Risk evaluation and long-term prediction of contaminant propagation

First step: Understanding the relevant processes



describing the relevant domains



and identifying the relevant scales

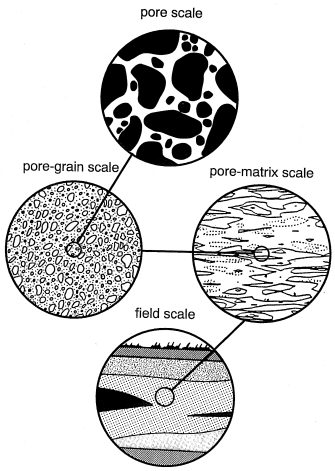


Figure 3.2: Spatial scales for modeling of groundwater flow and transport (Spitz & Moreno 1996).

Porous medium is

- **microscopically** a discontinuous medium consisting of solid matrix and pores
- **macroscopically** a continuum with spatially averaged effective characteristics
  Example: Porosity = total pore volume / volume of REV (representative elementary volume)
- Everything is a question of scale being considered!

3.1.1 Modeling flow in the subsurface

Darcy's Law

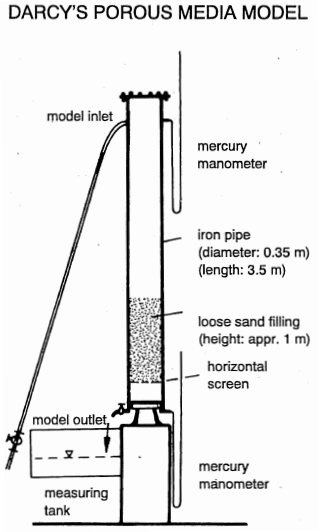


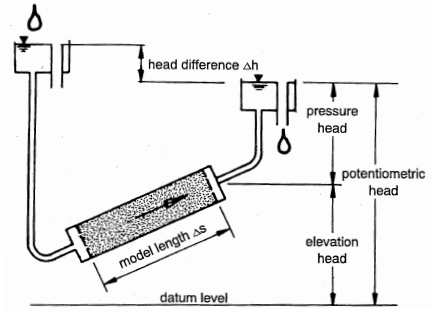
Figure 3.3: Experimental setup for Darcy's law (1856)

Observation: Water flux is proportional to the pressure gradient: $q = -K\dfrac{\Delta h}{\Delta s}\left(= \dfrac{Q}{A}\right)$

q: filtering velocity/volumetric flux density of water (Darcy velocity) [L/T]
K: (saturated) hydraulic conductivity [L/T]
Δh: pressure difference divided by ρg [L]
Δs: length of the column [L]
Q: flux (water volume per unit of time) [L^3/T]
A: (constant) cross-section of the column [L^2]

Remark 3.1. *Darcy's law is valid for stationary laminar flow in isotropic homogeneous porous media*

The total mechanical energy per unit of weight of water is nearly equal to its potential energy (due to very small flow velocities); the latter generally consists of the pressure and elevation parts.



The **piezometric height** also known as **piezometric or potentiometric head** h results from the action of the fluid pressure p and of the local height z

$$h = \frac{p}{\rho g} + z$$

h can be directly measured in the field using an observation well. The piezometric head is the state unknown of Darcy's equation.

Remark 3.2. *Darcy's law can be derived from the Stokes equations under the assumptions of stationary Newtonian flow of viscous fluid.*

Observation (Hubbert [1856]): Hydraulic conductivity K depends on

- the porous medium ($\rightarrow$ permeability $k[L^2]$)
- physical properties of the fluid ($\rightarrow$ density $\rho[M/L^3]$ and dynamic viscosity $\mu[M/(LT)]$)

and is given by $K = k\frac{\rho g}{\mu}$.

The hydraulic conductivity can be anisotropic, thus we obtain in the 3D

$$\boldsymbol{q} = \begin{pmatrix} q_x \\ q_y \\ q_z \end{pmatrix} = -\boldsymbol{K}\nabla h = -\begin{pmatrix} K_{xx}K_{xy}K_{xz} \\ K_{yx}K_{yy}K_{yz} \\ K_{zx}K_{zy}K_{zz} \end{pmatrix} \begin{pmatrix} \frac{\partial h}{\partial x} \\ \frac{\partial h}{\partial y} \\ \frac{\partial h}{\partial z} \end{pmatrix}$$

In an inhomogeneous case, K is, in addition, a function of the space coordinates: $\boldsymbol{K} = \boldsymbol{K}(x, y, z)$.

Saturated case
Conservation of mass (continuity equation)

- Mass- or volume conservation for water:

$$\nabla \cdot \boldsymbol{q} = \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} = 0.$$

- Generalization: Change in storage (S – storativity of porous medium)

$$\nabla \cdot \boldsymbol{q} = \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} = S\frac{\partial h}{\partial t}.$$

- Generalization: Sources/sinks (Q – volumetric density of sources/sinks)

$$\nabla \cdot \boldsymbol{q} = \frac{\partial q_x}{\partial x} + \frac{\partial q_y}{\partial y} + \frac{\partial q_z}{\partial z} = S\frac{\partial h}{\partial t} + Q.$$

Boundary conditions

- $h(t, \boldsymbol{x}) = g_1 \quad \text{on} \quad (0, T] \times \Gamma_1$ – DIRICHLET: Specified piezometric head h e.g. measured water levels, open water bodies

- $\boldsymbol{q}(t,\boldsymbol{x})\cdot\boldsymbol{n} = g_2 \quad \text{on} \quad (0,T]\times\Gamma_2$ – NEUMANN): Specified flux over the boundary q e.g. pumping wells, impermeable boundaries
- $\boldsymbol{q}(t,\boldsymbol{x})\cdot\boldsymbol{n} = I(h-h_{out}) \quad \text{to} \quad (0,T]\times\Gamma_3$ – MIXED, leakage, semipermeable: Flux depends on the difference between the external and the internal head e.g. drainage, semi-permeable river bed

Unsaturated case: Richards equation

Idea: Introduce moisture content $\Theta(h_p)$ (Volume of water/total volume); Seek: h_p and $\boldsymbol{q}$ in $(0,T]\times\Omega$ satisfying

$$\frac{\partial\Theta(h_p)}{\partial t} + \nabla\cdot\boldsymbol{q} = Q\,,$$
$$\boldsymbol{q} = -K_s\, k_r(h_p)\nabla(h_p+z)\,.$$

h_p:	pressure head ($h_p>0$ – saturated, $h_p<0$ – unsaturated)
$\boldsymbol{q}$:	volumetric velocity of water (Darcy velocity)
Q:	sources/sinks,
z:	height with respect to datum,
K_s:	saturated hydraulic conductivity, $K_s = \frac{k\rho g}{\mu}$ with permeability k, viscosity μ, density of water ρ, and the gravity constant g,
Θ:	volumetric moisture content $\Theta = \Theta(h_p)$,
k_r:	relative hydraulic conductivity $k_r = k_r(h_p)$.

Remark 3.3. Caution:

- ***Darcy velocity*** $\boldsymbol{q}$*: computed with the total REV*
- ***Average or seepage velocity*** $\boldsymbol{q}/\Theta$*: computed with the water-filled fraction of REV*

Generalizations of Darcy's law:

- Multi-phase flow
- Preferential flow directions
- Density-driven flows

- ...

Parametrization: Graph

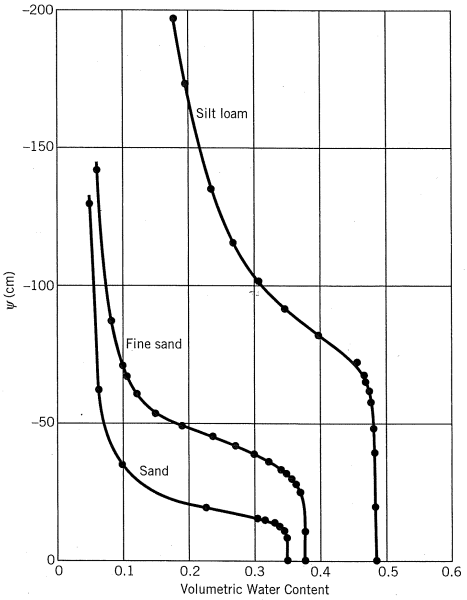


Figure 3.21 Water retention curves for sand, fine sand, and silt loam (from Brooks and Corey, 1966).

Typical non-linear pressure-saturation relationship (retention curve)

Parametrization: Analytical

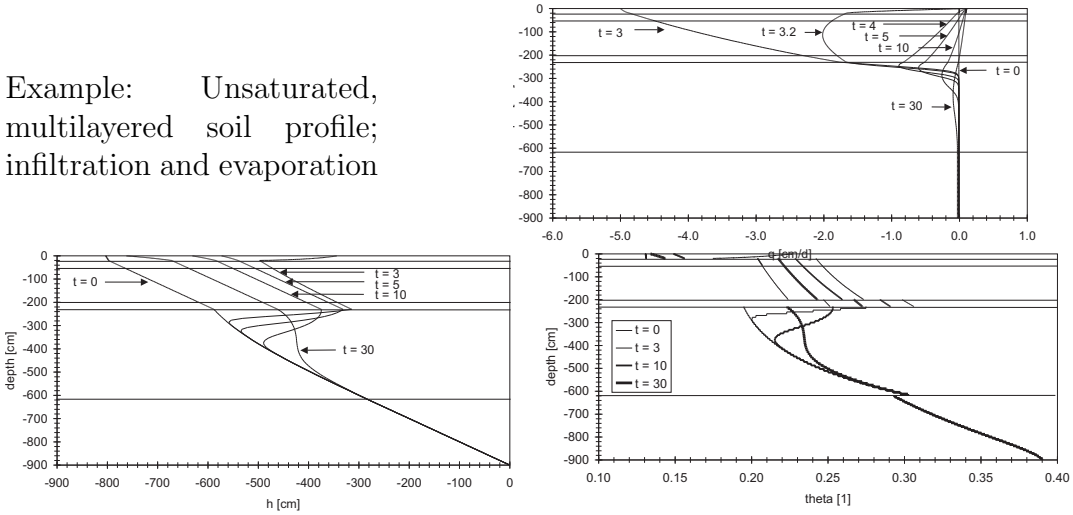
E.g. van Genuchten/Mualem

$$\Theta(h_p) = \Theta_r + (\Theta_s - \Theta_r)\left(\frac{1}{1+(-\alpha h_p)^m}\right)^{\frac{m}{m-1}}, \tag{3.1}$$

$$K(h_p) = K_s \frac{\left(1-(-\alpha h_p)^{m-1}(1+(-\alpha h_p)^m)^{\frac{1-m}{m}}\right)^2}{(1+(-\alpha h_p)^m)^{\frac{m-1}{2m}}}. \tag{3.2}$$

Need to specify: Θ_r, Θ_s, K_s, α, m.

Example: Unsaturated, multilayered soil profile; infiltration and evaporation



Questions

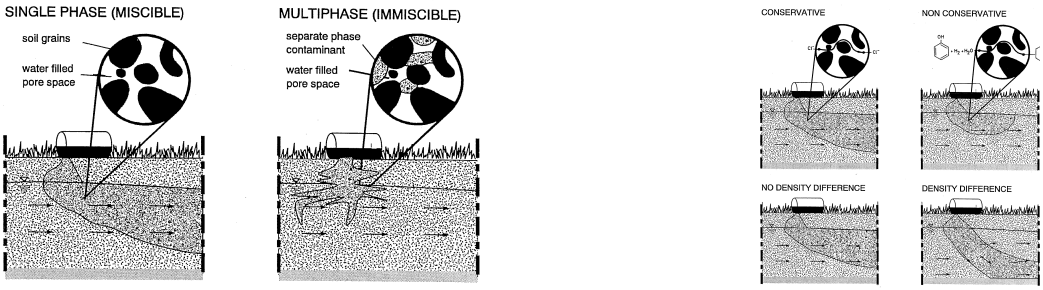
- Describe the application of the continuum approach to subsurface modeling at micro- and macro-scales.
- What is the main idea of Darcy's law and how can it be empirically confirmed?
- What is the piezometric head and how is it related to the pressure?
- How can the saturated flow model be extended to an unsaturated case?
- What methods can be used to determine the moisture content $\Theta(h_p)$, where h_p denotes the pressure head?

3.1.2 Modeling species transport in the subsurface

Species transport

Contaminants in the groundwater: Main problem settings

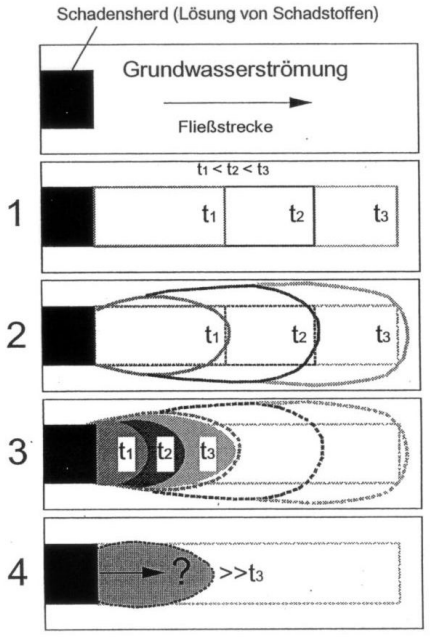
- Water quality: What is the natural groundwater background? Regional differences in the composition, different standards (US EPA, EU, ...)
- Contamination sources: Deponies, industry, agriculture, ...
- Thousands of substances are classified as dangerous
- Classification according to transport behavior



Main processes:

- *Advective transport*: Transport by the groundwater flow
- *Mechanical dispersion*: Mixing due to local variations in the velocity field
- *Molecular diffusion*: Mixing due to arbitrary movements of molecules (Brownian motion)
- *Adsorption*: Adhesion of fluid molecules to a surface
- *Degradation*: Destruction or transformation of species due to chemical, physical, or biological processes

Reactive species transport: Processes



Advection

Diffusion / Dispersion

Adsorption / Retention

Degradation / Transformation

Species transport model

- Specified: Description of groundwater flow
  $\Rightarrow$ $\Theta = \Theta(x,t)$ Moisture content
  $\boldsymbol{q} = \boldsymbol{q}(x,t)$ Darcy velocity

- Seek: $c = c(x,t)$ $[M/L^3]$: *Species concentration* in the water-filled part of REV)

- Main principle: Conservation of mass for the species Temporal change of concentration is equal to the flux through the domain boundaries plus the effect of sources/sinks

- Equation of mass conservation

$$\frac{\partial(\Theta c)}{\partial t} + \nabla \cdot \boldsymbol{J} = Q$$

$\boldsymbol{J} = \boldsymbol{J}(x,t)\,[M/L^2/T]$: *Mass flux*

$Q = Q(x,t)\,[M/L^3/T]$: *Sources/sinks*

Advection

- Transport of dissolved substances by the movement of solute

$$J = \Theta \boldsymbol{q} c$$

- Does not change the shape of the concentration front
- Frequently takes place in a complex setting:
  - unsaturated
  - heterogeneous media with preferred flow directions
  - cracks
  - density driven flow

Molecular diffusion
Material-dependent property

- Mass flux is proportional to the gradient of the concentration (Fick's Law)

$$J = -\Theta D_{\text{mol}} \nabla c$$

- Temperature-dependent
- Usually several orders of magnitude smaller than the dispersion or advection (there exist a few exceptions)
- Effective diffusion coefficient $D\,[L^2/T]$ in porous media is lower than the molecular diffusion coefficient in open water due to tortuosity of the media
- Causes species transport also without moving water!
- Irreversible process

Mechanical dispersion
Possible reasons:

- Microscopic heterogeneities in the porous medium:
  - Distribution of pore sizes
  - Varying pore geometry
- Macroscopic heterogeneties in the porous medium:
  - Layering
  - Differences in the permeability
  - Anisotropy in the geological structure of the domain
- Causes spreading of the contaminant
- Scale-depending process
- Analogy to diffusion: Fick's law
- Anisotropy (longitudinal / transversal dispersion differ!)

$$J = -\Theta \mathbf{D}_{\text{mech}} \nabla c$$
$$\mathbf{D}_{\text{mech}} = \mathbf{D}_{\text{mech}}(\boldsymbol{v}), \quad \boldsymbol{v} := \boldsymbol{q}/\Theta$$

Non-reactive one-component species transport
Species transport equation

$$\frac{\partial(\Theta c)}{\partial t} - \nabla \cdot (\Theta \boldsymbol{D} \nabla c - \Theta \boldsymbol{q} c) = 0.$$

with $\boldsymbol{D} = D_{\text{mol}} + \boldsymbol{D}_{\text{mech}}$ Generalization: Sources/sinks term Q

$$\frac{\partial(\Theta c)}{\partial t} - \nabla \cdot (\Theta \boldsymbol{D} \nabla c - \Theta \boldsymbol{q} c) = Q.$$

Boundary conditions

- DIRICHLET: $c(t, \boldsymbol{x}) = g_1 \quad \text{on} \quad (0, T] \times \Gamma_1$

- NEUMANN: $(D\nabla c) \cdot \boldsymbol{n} = g_2 \quad \text{on} \quad (0,T] \times \Gamma_2$
- MIXED: $\boldsymbol{q_3} c_3 = -D\nabla c + qc \quad \text{on} \quad (0,T] \times \Gamma_3$
- *Retention* by means of adsorption (but possibly
  - non-equilibrium,
  - mobile, reactive sorbents, ...)
- *Degradation* (but possibly
  - not at constant rates,
  - microbially catalyzed
  - temperature-dependent, ...)

Retention by means of adsorption

- Adsorption or sorption is the process of attachment of molecules of fluid or dissolved solid to solid matrix Adsorption substantially affects the mobility and spreading velocity!
- The mechanism of adsorption can be chemical (reactions) or/and physical (e.g. electric forces)
- Adsorption is not a homogeneous reaction (i.e. within a fluid phase) but rather inhomogeneous (fluid-solid)
- The rate of adsorption is influenced by the species concentration and properties, composition of the solid matrix, pH values, etc. (and is generally time- and space-dependent)
- The rate of adsorption may be (compared to the transport) fast ($\Rightarrow$ equilibrium assumption) or slow ($\Rightarrow$ kinetics)

Equilibrium adsorption

Definition 3.4. Adsorption isotherm *is a function describing the relationship between the* **dissolved concentration** $c = c(x,t) \quad [M/L^3]$ *and the* **adsorbed concentration** $s = s(x,t) \quad [M/M]$ *of the species at fixed temperature and under conditions of chemical equilibrium*

$$s(x,t) = f(c(x,t)).$$

- Adsorbed concentration is spatially **immobile**
- The measurement of adsorbed concentration is conducted in a laboratory by means of batch-tests: Solutions in different concentrations are mixed with the soil, and the residual species concentration in the fluid is measured
- The simplest relation (particularly useful at low concentration) is a linear one:
$$s = K_d c,$$
where $K_d \quad [L^3/M]$ denotes a distribution coefficient
- Non-linear **Freundlich** isotherm with exponent N:
$$s = K_d c^N$$
- both are empiric functions assuming that the solid matrix has an unlimited adsorbing capacity $\Rightarrow$ when fitting experimental data, no extrapolation in unmeasured ranges is allowed!
- **Langmuir** isotherm uses a concept of a finite number of attachment sites on solid matrix
- Once the maximum capacity $S_{\max}\,[M/M]$ is reached, no adsorption takes place:
$$s = \frac{K_d c}{1 + \frac{K_d}{S_{\max}} c}$$

Implications for the transport equation

$$\frac{\partial(\Theta c)}{\partial t} - \nabla \cdot (\Theta \boldsymbol{D} \nabla c - \Theta \boldsymbol{q} c) = -\rho_b \frac{\partial f(c)}{\partial t}.$$

with **deposit density** $\rho_b[M/L^3]$ (total mass/total volume) and the adsorption isotherm f.
Adsorption inhibits and slows down the transport.

In the linear case $f(c) = K_d c$, and constant Θ and ρ_b, one can describe the process using a constant **retardation** coefficient R:

$$\frac{\partial(\Theta c)}{\partial t} + \rho_b \frac{\partial(K_d c)}{\partial t} = \Theta \frac{\partial c}{\partial t} + \rho_b K_d \frac{\partial c}{\partial t} = (\Theta + \rho_b K_d) \frac{\partial c}{\partial t}$$

Dividing by Θ yields the retardation coefficient $R = 1 + \frac{\rho_b K_d}{\Theta}$.

Non-equilibrium adsorption

Definition 3.5. *Kinetics is the process of adsorption that is slow compared to transport; in this case the equilibrium conditions cannot be reached and a so-called kinetic adsorption model must be used*

$$\frac{\partial(\Theta c)}{\partial t} - \nabla \cdot (\Theta \boldsymbol{D} \nabla c - \Theta \boldsymbol{q} c) = -\rho_b \frac{\partial s}{\partial t}.$$

As the simplest model for adsorption kinetic one can take e.g.

$$\frac{\partial s}{\partial t} = k(f(c) - s)$$

with the rate factor $k[1/T]$.

Biodegradation

What are the main processes for the microbiological attenuation of organic contaminants?

- Redox reactions (chemical reactions resulting in electron exchanges between atoms) influence the concentration distributions of donors an acceptors of electrons
- Contaminant concentrations

- Biomass concentration and activity
- Inhibiting factors (temperature, toxicity)
- ...

0th order model

$$\frac{\partial(\Theta c)}{\partial t} - \nabla \cdot (\Theta \boldsymbol{D} \nabla c - \Theta \boldsymbol{q} c) = -R$$

constant loss of a fixed amount R.
It can be used in special cases as a simplification of the Monod model, e.g. if $c_D >> K_D$, constant activity of the biomass, e.g. due to limited availability of nutrients

1st order model

$$\frac{\partial(\Theta c)}{\partial t} - \nabla \cdot (\Theta \boldsymbol{D} \nabla c - \Theta \boldsymbol{q} c) = -kc$$

or more general as a reaction network

$$\frac{\partial(\Theta c_i)}{\partial t} - \nabla \cdot (\Theta \boldsymbol{D} \nabla c_i - \Theta \boldsymbol{q} c_i) = y c_{i-1} - k c_i$$

Destruction of the first order with rate k [1/T], parent species concentration c_{i-1} and **stoichiometric** coefficient y.
It can be used as an approximation if $c_D << K_D$, and constant biomass activity, e.g. due to limited availability of nutrients

Degradation using Monod kinetics

Simplest form of an one-component model: Assumption of constant biomass concentration, no limiting due to reaction partners:

$$\frac{\partial(\Theta c)}{\partial t} - \nabla \cdot (\Theta \boldsymbol{D} \nabla c - \Theta \boldsymbol{q} c) = -\Theta \mu_{\max} c_X \frac{c}{K + c}$$

with the maximum growth rate $\mu_{\max}[1/T]$, biomass concentration c_X, and the half-saturation constant $K[M/L^3]$.
This kinetics is also known as the **Michaelis-Menten** kinetics.

Monod kinetics

Monod kinetics even more general:

Reaction rate arbitrary electron donators, acceptors, inhibitors, and dynamic biomass:

$$\begin{aligned} R_r &= R_r(c_1, \ldots, c_{N_S}, c_{X_r}) \\ &= \mu_{\max_r} c_{X_r} \prod_{i \in I_r \subset \{1,\ldots,N_S\}} \left(\frac{c_i}{K_{M_i} + c_i} \right) \prod_{j \in J_r \subset \{1,\ldots,N_S\}} \left(\frac{K_{I_j}}{K_{I_j} + c_j} \right). \end{aligned}$$

Benzene remediation example

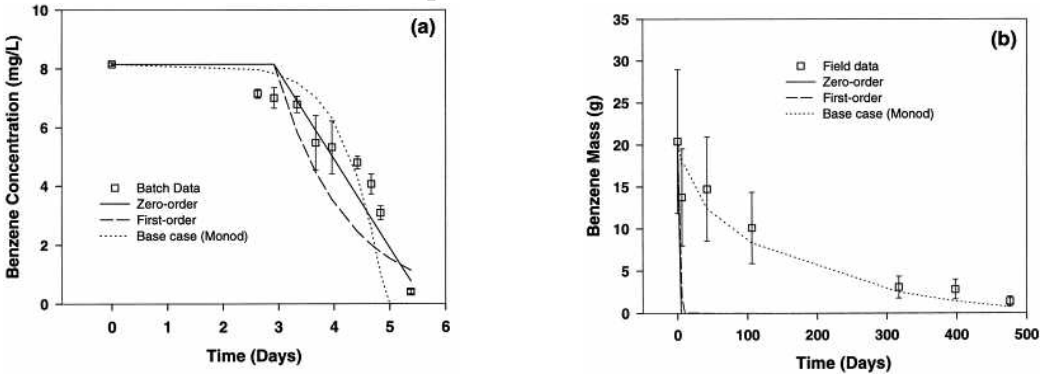


Figure 3.4: Lab experiments for determining the degradation rates for BTEX (Schirmer et al. 2000).

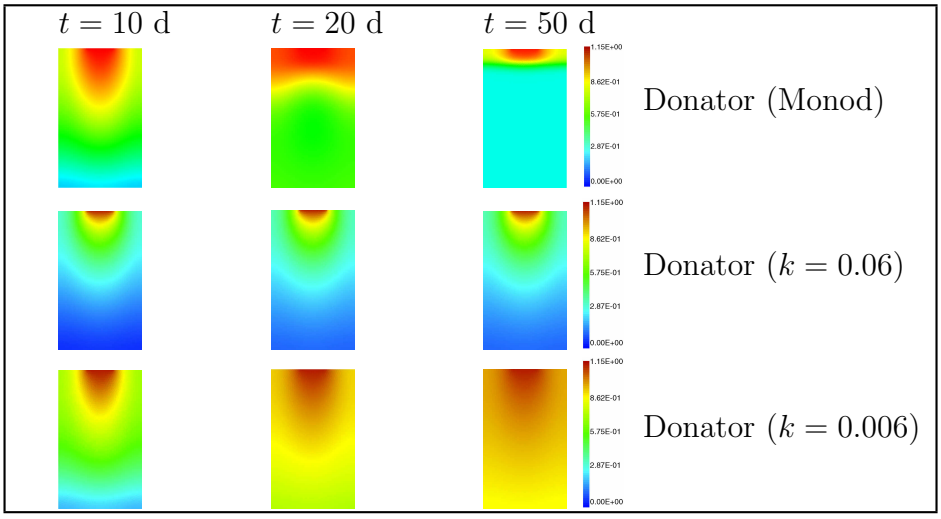


Figure 3.5: Simulation of benzene remediation (Schirmer et al. 2000).