Balance in the Atmosphere: Implications for Numerical Weather Prediction

Peter Lynch
School of Mathematical Sciences
University College Dublin

BIRS Summer School, Banff, 10–15 July, 2011

Outline

Introduction to Initialization

Richardson's Forecast

Scale Analysis of the SWE [Skip]

Early Initialization Methods

Laplace Tidal Equations [Skip]

Normal Mode Initialization

The Swinging Spring [Skip]

Digital Filter Initialization

Outline

Introduction to Initialization

- The spectrum of atmospheric motions is vast, encompassing phenomena having periods ranging from seconds to millennia.
- The motions of primary interest have relatively long timescales.
- The mathematical models used for numerical prediction describe a broader span of dynamical features than those of direct concern.

- For many purposes, the higher frequency (HF) components can be regarded as **noise**.
- The elimination of this noise is achieved by adjustment of the initial fields.
- This process is called **initialization**.

$\star \quad \star \quad \star$

Outline

Richardson’s Forecast

- The story of Lewis Fry Richardson’s forecast is well known.
- Richardson forecast the change in surface pressure at a point in central Europe, using the mathematical equations.

Richardson's Forecast

- The story of Lewis Fry Richardson's forecast is well known.
- Richardson forecast the change in surface pressure at a point in central Europe, using the mathematical equations.
- His results implied a change in surface pressure of 145 hPa in 6 hours ! ! !
- As Sir Napier Shaw remarked, "the wildest guess . . . would not have been wider of the mark . . . ".

- Yet, Richardson claimed that his forecast was *". . . a fairly correct deduction from a somewhat unnatural initial distribution".*
- He ascribed the unrealistic value of pressure tendency to errors in the observed winds.
- This is only a partial explanation of the problem.

The Spectrum of Atmospheric Motions

Atmospheric oscillations fall into two groups:

- Rotational or vortical modes (RH waves)
- Gravity-inertia wave oscillations

For typical conditions of large scale atmospheric flow the two types of motion are clearly separated and interactions between them are weak.

The Spectrum of Atmospheric Motions

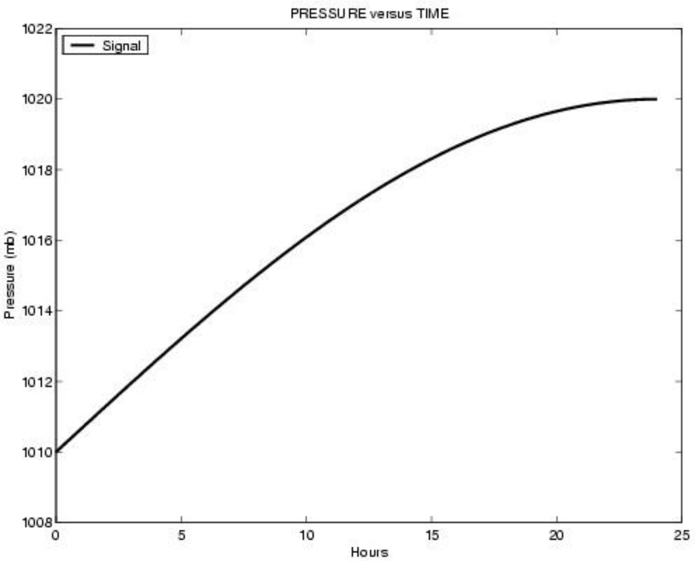
Atmospheric oscillations fall into two groups:

- Rotational or vortical modes (RH waves)
- Gravity-inertia wave oscillations

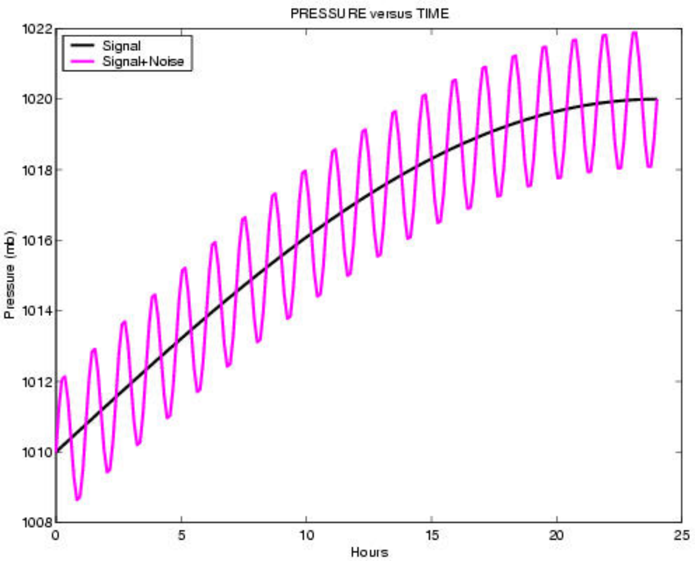
For typical conditions of large scale atmospheric flow the two types of motion are clearly separated and interactions between them are weak.

The high frequency gravity-inertia waves may be locally significant in the vicinity of steep orography, where there is strong thermal forcing or where very rapid changes are occurring . . .
. . . but overall they are of minor importance and may be regarded as undesirable noise.

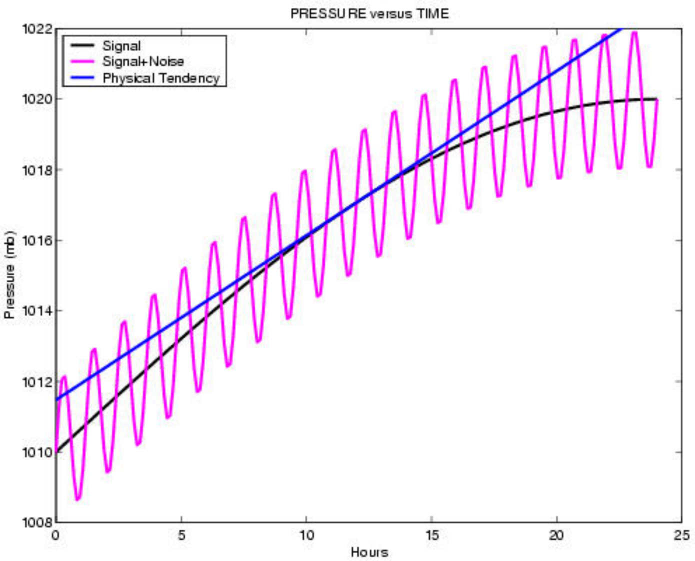
Smooth Evolution of Pressure



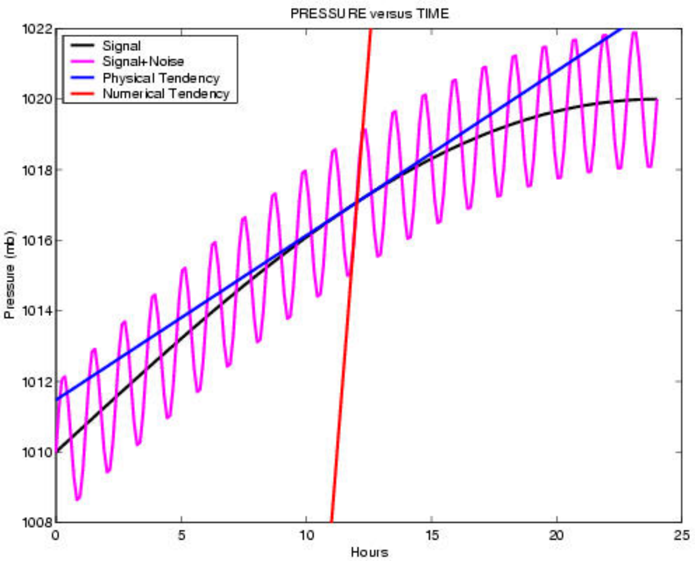
Noisy Evolution of Pressure



Tendency of a Smooth Signal



Tendency of a Noisy Signal



A Richardsonian Limerick

Young Richardson wanted to know
How quickly the pressure would grow.
But, what a surprise, ’cos
The six-hourly rise was,
In Pascals, One Four Five — Oh Oh!

A Simple Example — Ocean Tides

Imagine that you are standing by the sea shore on a stormy day.

The tidal variation, the slow changes between low and high water, has a period of about twelve hours.

A Simple Example — Ocean Tides

Imagine that you are standing by the sea shore on a stormy day.

The tidal variation, the slow changes between low and high water, has a period of about twelve hours.

Water level changes due to sea and swell have periods of less than a minute.

Clearly, the instantaneous value of water level cannot be used for tidal analysis.

A short time-scale wave.

At an instant, the water may be rising at a rate of one metre per second.

If the vertical velocity observed at an instant is used to predict the long-term movement of the water, a nonsensical forecast is obtained:

$$\text{Rise rate} = 3{,}600\,\text{m/hr} > 20\text{ km in 6 hours}$$

At an instant, the water may be rising at a rate of one metre per second.

If the vertical velocity observed at an instant is used to predict the long-term movement of the water, a nonsensical forecast is obtained:

Rise rate = 3,600 m/hr > 20 km in 6 hours

The instantaneous rate-of-change is no a guide to the long-term evolution.

The same is true of the atmosphere!

The Problem of Initialization.

A subtle and delicate state of balance exists in the atmosphere between the wind and pressure fields.

The fast gravity waves have much smaller amplitude than the slow rotational part of the flow.

The pressure and wind fields in regions not too near the equator are close to a state of geostrophic balance and the flow is quasi-nondivergent.

The Problem of Initialization.

A subtle and delicate state of balance exists in the atmosphere between the wind and pressure fields.

The fast gravity waves have much smaller amplitude than the slow rotational part of the flow.

The pressure and wind fields in regions not too near the equator are close to a state of geostrophic balance and the flow is quasi-nondivergent.

The existence of this geostrophic balance is a perennial source of interest.

It is a consequence of the forcing mechanisms and dominant modes of hydrodynamic instability and of the manner in which energy is dispersed and dissipated in the atmosphere.

When the primitive equations are used for numerical prediction the forecast may contain spurious large amplitude high frequency oscillations.

These result from anomalously large gravity-inertia wave components, arising from in the observations and analysis.

When the primitive equations are used for numerical prediction the forecast may contain spurious large amplitude high frequency oscillations.

These result from anomalously large gravity-inertia wave components, arising from in the observations and analysis.

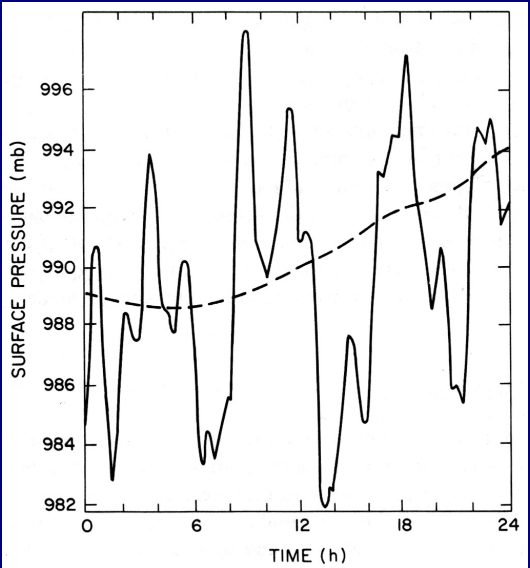
It was the presence of such imbalance in the initial fields that gave rise to the totally unrealistic pressure tendency of 145 hPa/6h obtained by Lewis Fry Richardson.

When the primitive equations are used for numerical prediction the forecast may contain spurious large amplitude high frequency oscillations.

These result from anomalously large gravity-inertia wave components, arising from in the observations and analysis.

It was the presence of such imbalance in the initial fields that gave rise to the totally unrealistic pressure tendency of 145 hPa/6h obtained by Lewis Fry Richardson.

The problems associated with high frequency motions are overcome by the process known as *initialization.*



Evolution of surface pressure before and after NNMI.
(Williamson and Temperton, 1981)

Need for Initialization

The principal aim of initialization is to define the initial fields so that the gravity inertia waves remain small throughout the forecast.

Need for Initialization

The principal aim of initialization is to define the initial fields so that the gravity inertia waves remain small throughout the forecast.

Specific requirements for initialization:

- Essential for satisfactory data assimilation
- Noisy forecasts have unrealistic vertical velocity
- Hopelessly inaccurate short-range rainfall patterns
- Spin-up of the humidity/water fields.
- Imbalance can lead to numerical instabilities.

Outline

Scale-analysis of the SWE

A scale analysis of the SWE is detailed on the following slides.

It will be omitted from the presentation, but is available for later study.

(Skip to next section.)

Scale-analysis of the SWE

We introduce **characteristic scales** for the dependent variables, and examine the relative sizes of the terms in the equations.

Scale-analysis of the SWE

We introduce **characteristic scales** for the dependent variables, and examine the relative sizes of the terms in the equations.

- **Length scale:** $\mathsf{L} = 10^6\,\mathbf{m}$
- **Velocity scale:** $\mathsf{V} = 10\,\mathbf{m\,s}^{-1}$
- **Advective time scale:** $\mathsf{T} = \mathsf{L}/\mathsf{V} = 10^5\,\mathbf{s}$
- **Pressure variation scale:** P
- **Scale height:** $\mathsf{H} = 10^4\,\mathbf{m}$
- **Acceleration of gravity:** $g = 10\,\mathbf{m\,s}^{-2}$
- **Coriolis parameter:** $f = 10^{-4}\,\mathbf{s}^{-1}$
- **Density:** $\rho_0 = 1\,\mathbf{kg\,m}^{-3}$

Scale-analysis of the SWE

We introduce characteristic scales for the dependent variables, and examine the relative sizes of the terms in the equations.

- **Length scale:** $\mathrm{L} = 10^6\,\mathrm{m}$
- **Velocity scale:** $\mathrm{V} = 10\,\mathrm{m\,s^{-1}}$
- **Advective time scale:** $\mathrm{T} = \mathrm{L}/\mathrm{V} = 10^5\,\mathrm{s}$
- **Pressure variation scale:** P
- **Scale height:** $\mathrm{H} = 10^4\,\mathrm{m}$
- **Acceleration of gravity:** $g = 10\,\mathrm{m\,s^{-2}}$
- **Coriolis parameter:** $f = 10^{-4}\,\mathrm{s^{-1}}$
- **Density:** $\rho_0 = 1\,\mathrm{kg\,m^{-3}}$

For simplicity, we may assume that $\rho_0 \equiv 1$.

The linear rotational **shallow water equations** are:

$$\underbrace{\frac{\partial u}{\partial t}}_{V^2/L} - \underbrace{fv}_{f\,V} + \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial x}}_{P/L} = 0$$

$$\underbrace{\frac{\partial v}{\partial t}}_{V^2/L} + \underbrace{fu}_{f\,V} + \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial y}}_{P/L} = 0$$

$$\underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial t}}_{PV/L} + \underbrace{gH\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right]}_{gHV/L} = 0$$

The scale of each term in the equations is indicated.

The linear rotational **shallow water equations** are:

$$\underbrace{\frac{\partial u}{\partial t}}_{\mathsf{V^2/L}} - \underbrace{fv}_{\mathsf{f\,V}} + \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial x}}_{\mathsf{P/L}} = 0$$

$$\underbrace{\frac{\partial v}{\partial t}}_{\mathsf{V^2/L}} + \underbrace{fu}_{\mathsf{f\,V}} + \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial y}}_{\mathsf{P/L}} = 0$$

$$\underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial t}}_{\mathsf{PV/L}} + \underbrace{gH\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right]}_{\mathsf{gHV/L}} = 0$$

The scale of each term in the equations is indicated.

If there is approximate balance between the Coriolis and pressure gradient terms, we must have

$$\frac{\mathsf{P}}{\mathsf{L}} = f\mathsf{V} \qquad \text{or} \qquad \mathsf{P} = f\mathsf{LV} = 10^3\ \textbf{Pa}$$

The ratio of the velocity tendencies to the Coriolis terms is the **Rossby number**

$$\mathrm{Ro} \equiv \frac{V}{fL} = \frac{10}{10^{-4} \cdot 10^{6}} = 10^{-1}\,,$$

a small parameter.

The ratio of the velocity tendencies to the Coriolis terms is the **Rossby number**

$$\mathrm{Ro} \equiv \frac{V}{fL} = \frac{10}{10^{-4}\cdot 10^{6}} = 10^{-1}\,,$$

a small parameter.

The scales of terms in the momentum equations are

$$\underbrace{\frac{\partial u}{\partial t}}_{10^{-4}} \quad - \quad \underbrace{fv}_{10^{-3}} \quad + \quad \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial x}}_{10^{-3}} \quad = \quad 0$$

$$\underbrace{\frac{\partial v}{\partial t}}_{10^{-4}} \quad + \quad \underbrace{fu}_{10^{-3}} \quad + \quad \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial y}}_{10^{-3}} \quad = \quad 0\,.$$

The ratio of the velocity tendencies to the Coriolis terms is the **Rossby number**

$$\mathrm{Ro} \equiv \frac{V}{fL} = \frac{10}{10^{-4} \cdot 10^{6}} = 10^{-1}\,,$$

a small parameter.

The scales of terms in the momentum equations are

$$\begin{aligned}
\underbrace{\frac{\partial u}{\partial t}}_{10^{-4}} &- \underbrace{fv}_{10^{-3}} + \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial x}}_{10^{-3}} = 0 \\
\underbrace{\frac{\partial v}{\partial t}}_{10^{-4}} &+ \underbrace{fu}_{10^{-3}} + \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial y}}_{10^{-3}} = 0\,.
\end{aligned}$$

To the lowest order of approximation, the tendency terms are negligible; there is **geostrophic balance** between the Coriolis and pressure terms.

Scaling the Divergence

Due to the cancellation between the two terms in the divergence, one might expect it to scale an order of magnitude smaller than each of its terms:

$$\delta = \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \sim \mathrm{Ro}\,\frac{\mathrm{V}}{\mathrm{L}} = 10^{-6}\,\mathrm{s}^{-1} \qquad \textbf{(?)}$$

Scaling the Divergence

Due to the cancellation between the two terms in the divergence, one might expect it to scale an order of magnitude smaller than each of its terms:

$$\delta = \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \sim \mathsf{Ro}\,\frac{\mathsf{V}}{\mathsf{L}} = 10^{-6}\,\mathrm{s}^{-1} \quad \textbf{(?)}$$

Then if we take $g = 10\,\mathbf{m\,s}^{-2}$ **and** $\mathsf{H} = 10^4$ **m, the continuity equation scales as**

$$\underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial t}}_{10^{-2}} \quad + \quad \underbrace{g\mathsf{H}\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right]}_{10^{-1}} \quad = \quad 0 \quad \textbf{???}$$

Scaling the Divergence

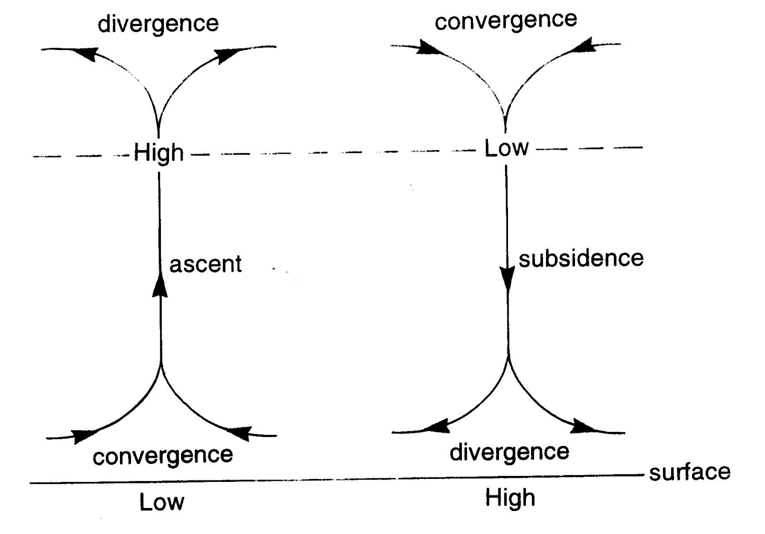
Due to the cancellation between the two terms in the divergence, one might expect it to scale an order of magnitude smaller than each of its terms:

$$\delta = \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) \sim \mathrm{Ro}\,\frac{\mathrm{V}}{\mathrm{L}} = 10^{-6}\,\mathrm{s}^{-1} \quad \textbf{(?)}$$

Then if we take $g = 10\,\mathbf{m\,s}^{-2}$ and $\mathrm{H} = 10^4$ m, the continuity equation scales as

$$\underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial t}}_{10^{-2}} \quad + \quad \underbrace{g\mathrm{H}\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right]}_{10^{-1}} \quad = \quad 0 \quad \textbf{???}$$

Impossible: there is nothing to balance the second term.



**Dines Compensation mechanism:
Cancellation of converegence and divergence.**

We recall that the divergence term

$$g \int \delta \, dz \quad \approx \quad gH \left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right] .$$

arises through **vertical integration.**

We recall that the divergence term

$$g \int \delta \, dz \quad \approx \quad gH \left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right] .$$

arises through vertical integration.

There is a tendency for cancellation between convergence at low levels and divergence at higher levels. This is called the Dines compensation mechanism.

We recall that the divergence term

$$g \int \delta \, dz \quad \approx \quad gH \left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right] .$$

arises through **vertical integration.**

There is a tendency for **cancellation between convergence at low levels and divergence at higher levels.** This is called the **Dines compensation mechanism.**

Thus, we assume

$$\int \delta \, dz \sim \mathrm{Ro}\, \delta H \,, \qquad \text{so that} \qquad g \int \delta \, dz \sim \mathrm{Ro}^2 gH \frac{V}{L} = 10^{-2} \,.$$

The terms of the continuity equation are now in balance:

$$\underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial t}}_{10^{-2}} + \underbrace{gH\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right]}_{10^{-2}} = 0$$

The terms of the continuity equation are now in balance:

$$\underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial t}}_{10^{-2}} \quad + \quad \underbrace{gH\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right]}_{10^{-2}} \quad = \quad 0$$

So, $\partial p/\partial t \sim 10^{-2}$ Pa s^{-1}, which is about 1 hPa per 3 hours.

(Illustrate the Dines compensation mechanism for a cyclone.)

The Effect of Data Errors

Suppose there is a 10% error Δv in the v-component of the wind observation at a point.

The Effect of Data Errors

Suppose there is a 10% error Δv in the v-component of the wind observation at a point.

The scales of the terms are as before:

$$\underbrace{\frac{\partial u}{\partial t}}_{10^{-4}} \quad - \quad \underbrace{f(v + \Delta v)}_{10^{-3}} \quad + \quad \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial x}}_{10^{-3}} \quad = \quad 0$$

The Effect of Data Errors

Suppose there is a 10% error Δv in the v-component of the wind observation at a point.

The scales of the terms are as before:

$$\underbrace{\frac{\partial u}{\partial t}}_{10^{-4}} \quad - \quad \underbrace{f(v + \Delta v)}_{10^{-3}} \quad + \quad \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial x}}_{10^{-3}} \quad = \quad 0$$

However, the error in the tendency is $\Delta(\partial u/\partial t) \sim f\Delta v \sim 10^{-4}$, comparable in size to the tendency itself.

The signal-to-noise ratio is $\sim$1.

The Effect of Data Errors

Suppose there is a 10% error Δv in the v-component of the wind observation at a point.

The scales of the terms are as before:

$$\underbrace{\frac{\partial u}{\partial t}}_{10^{-4}} \quad - \quad \underbrace{f(v+\Delta v)}_{10^{-3}} \quad + \quad \underbrace{\frac{1}{\rho_0}\frac{\partial p}{\partial x}}_{10^{-3}} \quad = \quad 0$$

However, the error in the tendency is $\Delta(\partial u/\partial t) \sim f\Delta v \sim 10^{-4}$, comparable in size to the tendency itself.

The signal-to-noise ratio is $\sim$1.

The forecast may be qualitatively reasonable, but it will be **quantitatively invalid**.

A similar conclusion is reached for a 10% error in the pressure gradient.

A similar conclusion is reached for a 10% error in the pressure gradient.

However, if the spatial scale Δx of the pressure error is small (say, $\Delta x \sim \mathrm{L}/10$) the error in its gradient is correspondingly large:

$$\frac{\partial p}{\partial x} \sim \frac{\mathrm{P}}{\mathrm{L}}\,, \qquad \text{but} \qquad \Delta\frac{\partial p}{\partial x} \sim \frac{\Delta p}{\Delta x} \sim \frac{\mathrm{P}}{\mathrm{L}} \sim \frac{\partial p}{\partial x}\,,$$

A similar conclusion is reached for a 10% error in the pressure gradient.

However, if the spatial scale Δx of the pressure error is small (say, $\Delta x \sim \mathrm{L}/10$) the error in its gradient is correspondingly large:

$$\frac{\partial p}{\partial x} \sim \frac{\mathrm{P}}{\mathrm{L}}, \qquad \text{but} \qquad \Delta\frac{\partial p}{\partial x} \sim \frac{\Delta p}{\Delta x} \sim \frac{\mathrm{P}}{\mathrm{L}} \sim \frac{\partial p}{\partial x},$$

Thus, that the error in the wind tendency is now

$$\Delta\frac{\partial u}{\partial t} \sim \frac{1}{\rho_0}\frac{\partial p}{\partial x} \sim 10^{-3} \gg \frac{\partial u}{\partial t}.$$

The forecast will be qualitatively incorrect (i.e., useless!).

Now consider the continuity equation. The pressure tendency has scale

$$\frac{\partial p}{\partial t} \sim 10^{-2}\,\mathrm{Pa\,s^{-1}} \approx 1\,\mathrm{hPa\ in\ 3\ hours}\,.$$

Now consider the **continuity equation**. The pressure tendency has scale

$$\frac{\partial p}{\partial t} \sim 10^{-2}\,\mathrm{Pa\,s}^{-1} \approx 1\,\mathrm{hPa\ in\ 3\ hours}\,.$$

If there is **a 10% error in the wind**, the resulting error in divergence is $\Delta\delta \sim \Delta v/L \sim 10^{-6}$.

Now consider the continuity equation. The pressure tendency has scale

$$\frac{\partial p}{\partial t} \sim 10^{-2}\,\mathrm{Pa\,s}^{-1} \approx 1\ \mathrm{hPa\ in\ 3\ hours}\,.$$

If there is a 10% error in the wind, the resulting error in divergence is $\Delta\delta \sim \Delta v/L \sim 10^{-6}$.

The error is larger than the divergence itself! Thus, the pressure tendency is unrealistic.

Now consider the continuity equation. The pressure tendency has scale

$$\frac{\partial p}{\partial t} \sim 10^{-2}\,\mathrm{Pa\,s^{-1}} \approx 1\ \mathrm{hPa\ in\ 3\ hours}\,.$$

If there is a 10% error in the wind, the resulting error in divergence is $\Delta\delta \sim \Delta v/L \sim 10^{-6}$.

The error is larger than the divergence itself! Thus, the pressure tendency is unrealistic.

Worse still, if the wind error is of small spatial scale, the divergence error is correspondingly greater:

$$\Delta\delta \sim \Delta\frac{\partial v}{\partial x} \sim \frac{\Delta v}{\Delta x} \sim \frac{V}{L} \sim 10^{-5} \sim 10^{2}\delta\,.$$

Now consider the continuity equation. The pressure tendency has scale

$$\frac{\partial p}{\partial t} \sim 10^{-2}\,\mathrm{Pa\,s^{-1}} \approx 1\,\mathrm{hPa\ in\ 3\ hours}\,.$$

If there is a 10% error in the wind, the resulting error in divergence is $\Delta\delta \sim \Delta v/L \sim 10^{-6}$.

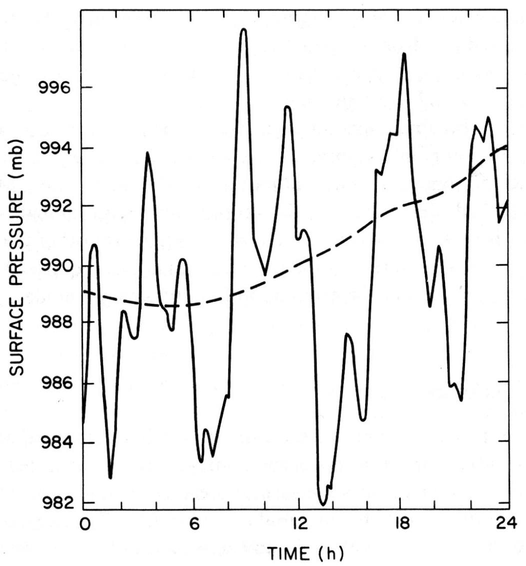
The error is larger than the divergence itself! Thus, the pressure tendency is unrealistic.

Worse still, if the wind error is of small spatial scale, the divergence error is correspondingly greater:

$$\Delta\delta \sim \Delta\frac{\partial v}{\partial x} \sim \frac{\Delta v}{\Delta x} \sim \frac{V}{L} \sim 10^{-5} \sim 10^{2}\delta\,.$$

This implies a pressure tendency two orders of magnitude larger than the correct value.

Instead of the value $\partial p/\partial t \sim 1$ hPa in 3 hours we get a change of order 100 hPa in 3 hours (like Richardson's result).



Evolution of surface pressure before and after NNMI.
(Williamson and Temperton, 1981)

Outline

Early Initialization Methods

We will describe, in outline, a number of methods that have been used to overcome the problems of noise in numerical integrations.

- 1. The Filtered Equations
- 2. Static Initialization
- 3. Dynamic Initialization
- 4. Variational Initialization

1. The Filtered Equations

The first computer forecast was made in 1950 by Charney, Fjørtoft and Von Neumann, using

$$\frac{d}{dt}(\zeta + f) = 0$$

which has no gravity wave components.

Systems like this are called Filtered Equations. The basic filtered system is the QG equations.

1. The Filtered Equations

The first computer forecast was made in 1950 by Charney, Fjørtoft and Von Neumann, using

$$\frac{d}{dt}(\zeta + f) = 0$$

which has no gravity wave components.

Systems like this are called Filtered Equations. The basic filtered system is the QG equations.

The *barotropic, quasi-geostrophic potential vorticity equation* (the QGPV Equation) is

$$\frac{\partial}{\partial t}\left(\nabla^2\psi - F\psi\right) + J(\psi, \nabla^2\psi) + \beta\frac{\partial\psi}{\partial x} = 0\,.$$

This is a *single equation* for a *single variable*, ψ.

The simplifying assumptions have the effect of **eliminating high-frequency gravity wave solutions**, so that only the slow Rossby wave solutions remain.

⋆ ⋆ ⋆

The simplifying assumptions have the effect of eliminating high-frequency gravity wave solutions, so that only the slow Rossby wave solutions remain.

$\star \quad \star \quad \star$

A more accurate filtering of the primitive equations leads to the balance equations.

This system is more complicated to solve than the QG system. It has not been widely used.

However one diagnostic component has been used for initialization. We discuss this presently.

2. Static Initialization

Hinkelmann (1951) investigated the problem of noise in numerical integrations of the primitive equations.

He concluded that, if initial winds were geostrophic,

$$\mathbf{V} = \frac{1}{f}\mathbf{k} \times \nabla\Phi \,,$$

HF oscillations would remain small in amplitude.

2. Static Initialization

Hinkelmann (1951) investigated the problem of noise in numerical integrations of the primitive equations.

He concluded that, if initial winds were geostrophic,

$$\mathbf{V} = \frac{1}{f}\mathbf{k} \times \nabla\Phi \,,$$

HF oscillations would remain small in amplitude.

If we express the wind as $\mathbf{V} = \mathbf{k} \times \nabla\psi$, we can write

$$f\nabla\psi = \nabla\Phi$$

The divergence of this is the linear balance equation:

$$\nabla\cdot f\nabla\psi = \nabla^2\Phi$$

This can be solved for ψ if Φ is given, or for Φ if ψ is given.

Forecasts made with the primitive equations were soon shown to be clearly superior to those using the quasi-geostrophic system . . .

. . . however, the use of geostrophic initial winds has a huge disadvantage:

Forecasts made with the primitive equations were soon shown to be clearly superior to those using the quasi-geostrophic system . . .

. . . however, the use of geostrophic initial winds has a huge disadvantage:

Observations of the wind field are completely ignored.

Forecasts made with the primitive equations were soon shown to be **clearly superior** to those using the quasi-geostrophic system . . .
. . . however, the use of geostrophic initial winds has a huge disadvantage:

Observations of the wind field are completely ignored.

Charney (1955) proposed a better estimate of the wind, using the **nonlinear balance equation.**

This equation is a diagnostic relationship between the pressure and wind fields.

$$\nabla^2\Phi - \nabla\cdot f\nabla\psi + 2\left[\left(\frac{\partial^2\psi}{\partial x\partial y}\right)^2 - \frac{\partial^2\psi}{\partial x^2}\frac{\partial^2\psi}{\partial y^2}\right] = 0$$

This is a Poisson equation for Φ when ψ is given. However, it is **nonlinear in ψ** and **hard to solve for ψ when Φ is given.**

When ψ is obtained from the nonlinear balance equation, a non-divergent wind is constructed:

$$\mathbf{V} = \mathbf{k} \times \nabla\psi\,.$$

Phillips (1960) argued that, in addition to getting ψ from the nonlinear balance equation, a divergent component of the wind should be included.

When ψ is obtained from the nonlinear balance equation, a non-divergent wind is constructed:

$$\mathbf{V} = \mathbf{k} \times \nabla\psi \,.$$

Phillips (1960) argued that, in addition to getting ψ from the nonlinear balance equation, a divergent component of the wind should be included.

He proposed that a further improvement would result if the divergence of the initial field were set equal to that implied by quasi-geostrophic theory.

This can be sone by solving the QG omega equation.

When ψ is obtained from the nonlinear balance equation, a non-divergent wind is constructed:

$$\mathbf{V} = \mathbf{k} \times \nabla\psi \,.$$

Phillips (1960) argued that, in addition to getting ψ from the nonlinear balance equation, a divergent component of the wind should be included.

He proposed that a further improvement would result if the divergence of the initial field were set equal to that implied by quasi-geostrophic theory.

This can be sone by solving the QG omega equation.

Each of these steps represented some progress, but the noise problem still remained essentially unsolved.

3. Dynamic Initialization

Another approach, called dynamic initialization, uses the forecast model itself to define the initial fields.

The dissipative processes in the model can damp out high frequency noise as the forecast procedes.

3. Dynamic Initialization

Another approach, called dynamic initialization, uses the forecast model itself to define the initial fields.

The dissipative processes in the model can damp out high frequency noise as the forecast procedes.

We integrate the model forward and backward in time, keeping the dissipation active all the time.

We repeat this forward-backward cycle many times until we obtain initial fields from which the high frequency components have been damped out.

The forecast starting from the dynamically balanced fields is noise-free . . .

. . . however, the procedure is expensive in terms of computer time.

The forecast starting from the dynamically balanced fields is noise-free . . .

. . . however, the procedure is expensive in terms of computer time.

Moreover, it damps the meteorologically significant motions as well as the gravity waves.

Thus, dynamic initialization is no longer popular.

3A. Digital filtering initialization (DFI)

Digital filtering initialization (DFI) is essentially a refinement of dynamic initialization.

Because it used a highly selective filtering technique, it is computationally more efficient than the older dynamic initialization method.

3A. Digital filtering initialization (DFI)

Digital filtering initialization (DFI) is essentially a refinement of dynamic initialization.

Because it used a highly selective filtering technique, it is computationally more efficient than the older dynamic initialization method.

If time permits, we will return to DFI later.

4. Variational Initialization

An elegant initialization method based on the calculus of variations was introduced by Sasaki (1958).

Although the method was not widely used, the variational method is now at the centre of modern data assimilation practice.

4. Variational Initialization

An elegant initialization method based on the calculus of variations was introduced by Sasaki (1958).

Although the method was not widely used, the variational method is now at the centre of modern data assimilation practice.

Recall that, in variational assimilation, we minimize a cost function, J, which is normally a sum of two terms

$$J = J_B + J_O$$

$$J(\mathbf{x}) = J_B(\mathbf{x}) + J_O(\mathbf{x})$$

Here, J_B is the distance between the analysis and the background field

$$J_B = \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)$$

and J_O is the distance to the observations

$$J_O = \tfrac{1}{2}[\mathbf{y}_o - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}_o - H(\mathbf{x}_b)]$$

The idea is to choose x so as to minimize $J(\mathbf{x})$.

$\star \quad \star \quad \star$

$$J(\mathbf{x}) = J_B(\mathbf{x}) + J_O(\mathbf{x})$$

Here, J_B is the distance between the analysis and the background field

$$J_B = \tfrac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b)$$

and J_O is the distance to the observations

$$J_O = \tfrac{1}{2}[\mathbf{y}_o - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}_o - H(\mathbf{x}_b)]$$

The idea is to choose x so as to minimize $J(\mathbf{x})$.

$\star \quad \star \quad \star$

The variational problem can be be modified to include a balance constraint.

We add a constraint which requires the analysis to be close to geostrophic balance:

$$J_C = \frac{1}{2}\alpha \sum_{ij} \left[\left(fu + \frac{\partial \Phi}{\partial y} \right)^2_{ij} + \left(fv - \frac{\partial \Phi}{\partial x} \right)^2_{ij} \right]$$

We add a **constraint** which requires the analysis to be close to geostrophic balance:

$$J_C = \tfrac{1}{2}\alpha \sum_{ij} \left[\left(fu + \frac{\partial \Phi}{\partial y} \right)_{ij}^2 + \left(fv - \frac{\partial \Phi}{\partial x} \right)_{ij}^2 \right]$$

This term J_C is large if the analysis is far from geostrophic balance. It **vanishes for perfect geostrophic balance.**

We add a constraint which requires the analysis to be close to geostrophic balance:

$$J_C = \frac{1}{2}\alpha \sum_{ij} \left[\left(fu + \frac{\partial \Phi}{\partial y} \right)_{ij}^2 + \left(fv - \frac{\partial \Phi}{\partial x} \right)_{ij}^2 \right]$$

This term J_C is large if the analysis is far from geostrophic balance. It vanishes for perfect geostrophic balance.

The weight α is chosen to give the constraint an appropriate impact. This is known as a weak constraint.

We add a **constraint** which requires the analysis to be close to geostrophic balance:

$$J_C = \tfrac{1}{2}\alpha \sum_{ij} \left[\left(fu + \frac{\partial\Phi}{\partial y} \right)^2_{ij} + \left(fv - \frac{\partial\Phi}{\partial x} \right)^2_{ij} \right]$$

This term J_C is large if the analysis is far from geostrophic balance. It **vanishes for perfect geostrophic balance**.

The weight α is chosen to give the constraint an appropriate impact. This is known as a **weak constraint**.

The constrained variational assimilation finds the minimum of the cost function

$$J = J_B + J_O + J_C$$

Outline

Atmospheric Normal Modes

The solutions of the atmospheric equations can be separated, by spectral analysis, into two sets of linear normal modes:

- Slow rotational components or Rossby modes
- High frequency gravity-inertia modes

Atmospheric Normal Modes

The solutions of the atmospheric equations can be separated, by spectral analysis, into two sets of linear normal modes:

- Slow rotational components or Rossby modes
- High frequency gravity-inertia modes

If the amplitude of the motion is small, the horizontal structure is then governed by a system equivalent to the linear shallow water equations.

Atmospheric Normal Modes

The solutions of the atmospheric equations can be separated, by spectral analysis, into two sets of linear normal modes:

- Slow rotational components or Rossby modes
- High frequency gravity-inertia modes

If the amplitude of the motion is small, the horizontal structure is then governed by a system equivalent to the linear shallow water equations.

These equations were first derived by Laplace in his discussion of tides in the atmosphere and ocean.

They are called the Laplace Tidal Equations.

The Laplace Tidal Equations

The simplest means of deriving the linear shallow water equations from the primitive equations is to assume that the vertical velocity vanishes identically.

The Laplace Tidal Equations

The simplest means of deriving the linear shallow water equations from the primitive equations is to assume that the vertical velocity vanishes identically.

We assume that the motions can be described as small perturbations about a state of rest, with constant temperature T_0, and pressure $\bar{p}(z)$ and density $\bar{\rho}(z)$ varying with height.

The basic state variables satisfy the gas law, and are in hydrostatic balance:

$$\bar{p} = \mathcal{R}\bar{\rho}T_0 \qquad \textbf{and} \qquad \frac{d\bar{p}}{dz} = -g\bar{\rho}$$

The basic state variables satisfy the gas law, and are in hydrostatic balance:

$$\bar{p} = \mathcal{R}\bar{\rho}T_0 \qquad \text{and} \qquad \frac{d\bar{p}}{dz} = -g\bar{\rho}$$

The variations of mean pressure and density follow:

$$\bar{p}(z) = p_0 \exp(-z/H) \,, \qquad \bar{\rho}(z) = \rho_0 \exp(-z/H) \,,$$

where $H = p_0/g\rho_0 = \mathcal{R}T_0/g$ is the atmospheric scale-height.

The **basic state variables** satisfy the gas law, and are in hydrostatic balance:

$$\bar{p} = \mathcal{R}\bar{\rho}T_0 \qquad \text{and} \qquad \frac{d\bar{p}}{dz} = -g\bar{\rho}$$

The variations of mean pressure and density follow:

$$\bar{p}(z) = p_0 \exp(-z/H)\,, \qquad \bar{\rho}(z) = \rho_0 \exp(-z/H)\,,$$

where $H = p_0/g\rho_0 = \mathcal{R}T_0/g$ is the atmospheric scale-height.

Exercise: Confirm this.

We consider only motions for which the vertical component of velocity vanishes identically, $w \equiv 0$.

We consider only motions for which the vertical component of velocity vanishes identically, $w \equiv 0$.

Let u, v, p and ρ denote variations about the basic state, each of these being a small quantity. Then

$$
\begin{aligned}
\frac{\partial \bar{\rho} u}{\partial t} - f\bar{\rho}v + \frac{\partial p}{\partial x} &= 0 \\
\frac{\partial \bar{\rho} v}{\partial t} + f\bar{\rho}u + \frac{\partial p}{\partial y} &= 0 \\
\frac{\partial \rho}{\partial t} + \nabla\cdot\bar{\rho}\mathbf{V} &= 0 \\
\frac{1}{\gamma\bar{p}}\frac{\partial p}{\partial t} - \frac{1}{\bar{\rho}}\frac{\partial \rho}{\partial t} &= 0
\end{aligned}
$$

We consider only motions for which the vertical component of velocity vanishes identically, $w \equiv 0$.

Let u, v, p and ρ denote variations about the basic state, each of these being a small quantity. Then

$$
\begin{aligned}
\frac{\partial \bar{\rho} u}{\partial t} - f\bar{\rho} v + \frac{\partial p}{\partial x} &= 0 \\
\frac{\partial \bar{\rho} v}{\partial t} + f\bar{\rho} u + \frac{\partial p}{\partial y} &= 0 \\
\frac{\partial \rho}{\partial t} + \nabla\cdot\bar{\rho}\mathbf{V} &= 0 \\
\frac{1}{\gamma\bar{p}}\frac{\partial p}{\partial t} - \frac{1}{\bar{\rho}}\frac{\partial \rho}{\partial t} &= 0
\end{aligned}
$$

Density can be eliminated from the continuity equation by means of the thermodynamic equation. We then get three equations for u, v and p.

We now assume that the horizontal and vertical dependencies of the perturbation quantities are separable:

$$\left\{\begin{matrix} \bar{\rho}u \\ \bar{\rho}v \\ p \end{matrix}\right\} = \left\{\begin{matrix} U(x,y,t) \\ V(x,y,t) \\ P(x,y,t) \end{matrix}\right\} Z(z) \,.$$

We now assume that the horizontal and vertical dependencies of the perturbation quantities are **separable**:

$$\left\{ \begin{array}{c} \bar{\rho}u \\ \bar{\rho}v \\ p \end{array} \right\} = \left\{ \begin{array}{c} U(x,y,t) \\ V(x,y,t) \\ P(x,y,t) \end{array} \right\} Z(z) \, .$$

The momentum and continuity equations become

$$\begin{aligned} \frac{\partial U}{\partial t} - fV + \frac{\partial P}{\partial x} &= 0 \\ \frac{\partial V}{\partial t} + fU + \frac{\partial P}{\partial y} &= 0 \\ \frac{\partial P}{\partial t} + (gh)\nabla \cdot \mathbf{V} &= 0 \end{aligned}$$

where $\mathbf{V} = (U, V)$ is the momentum and $h = \gamma H = \gamma \mathcal{R} T_0 / g$.

We now assume that the horizontal and vertical dependencies of the perturbation quantities are **separable**:

$$\begin{Bmatrix} \bar{\rho}u \\ \bar{\rho}v \\ p \end{Bmatrix} = \begin{Bmatrix} U(x,y,t) \\ V(x,y,t) \\ P(x,y,t) \end{Bmatrix} Z(z) \,.$$

The momentum and continuity equations become

$$\begin{aligned} \frac{\partial U}{\partial t} - fV + \frac{\partial P}{\partial x} &= 0 \\ \frac{\partial V}{\partial t} + fU + \frac{\partial P}{\partial y} &= 0 \\ \frac{\partial P}{\partial t} + (gh)\nabla\cdot\mathbf{V} &= 0 \end{aligned}$$

where $\mathbf{V} = (U, V)$ is the momentum and $h = \gamma H = \gamma\mathcal{R}T_0/g$.

This is a set of three equations for U, V, and P.

They are mathematically isomorphic to the **Laplace Tidal Equations** with a mean depth h (the **equivalent depth**).

The Vertical Structure Equation

The vertical structure follows from the hydrostatic equation, together with the relationship $p = (\gamma g H)\rho$ implied by the thermodynamic equation. It is determined by

$$\frac{dZ}{dz} + \frac{Z}{\gamma H} = 0 \ ,$$

The Vertical Structure Equation

The **vertical structure** follows from the hydrostatic equation, together with the relationship $p = (\gamma g H)\rho$ implied by the thermodynamic equation. It is determined by

$$\frac{dZ}{dz} + \frac{Z}{\gamma H} = 0 ,$$

The solution of this is $Z = Z_0 \exp(-z/\gamma H)$, where Z_0 is the amplitude at $z = 0$.

If we set $Z_0 = 1$, then U, V and P give the momentum and pressure fields at the earth's surface. These variables all decay exponentially with height.

It follows that u and v actually increase with height as $\exp(\kappa z/H)$, but the kinetic energy decays.

If we set $Z_0 = 1$, then U, V and P give the momentum and pressure fields at the earth's surface. These variables all decay exponentially with height.

It follows that u and v actually increase with height as $\exp(\kappa z/H)$, but the kinetic energy decays.

Solutions with more general vertical structures, and with non-vanishing vertical velocity, may be derived.

Vorticity and Divergence

We examine the solutions of the Laplace Tidal Equations in some enlightening limiting cases.

Vorticity and Divergence

We examine the solutions of the Laplace Tidal Equations in some enlightening limiting cases.

By means of the Helmholtz Theorem, a general horizontal wind field V may be partitioned into rotational and divergent components

$$\mathbf{V} = \mathbf{V}_\psi + \mathbf{V}_\chi = \mathbf{k} \times \nabla\psi + \nabla\chi \,.$$

Vorticity and Divergence

We examine the solutions of the Laplace Tidal Equations in some enlightening limiting cases.

By means of the Helmholtz Theorem, a general horizontal wind field V may be partitioned into rotational and divergent components

$$\mathbf{V} = \mathbf{V}_\psi + \mathbf{V}_\chi = \mathbf{k} \times \nabla\psi + \nabla\chi \,.$$

The stream function ψ and velocity potential χ are related to ζ and δ by the Poisson equations

$$\nabla^2\psi = \zeta \qquad \text{and} \qquad \nabla^2\chi = \delta \,.$$

By differentiating the momentum equations, we get equations for the vorticity and divergence tendencies, e.g.,

$$\frac{\partial \zeta}{\partial t} = \frac{\partial}{\partial t}\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) = \frac{\partial}{\partial x}\left(\frac{\partial v}{\partial t}\right) - \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial t}\right)$$

The vorticity, divergence and continuity equations are

$$
\begin{aligned}
\frac{\partial \zeta}{\partial t} + f\delta + \beta v &= 0 \\
\frac{\partial \delta}{\partial t} - f\zeta + \beta u + \nabla^2 P &= 0 \\
\frac{\partial P}{\partial t} + gh\delta &= 0\,.
\end{aligned}
$$

The vorticity, divergence and continuity equations are

$$
\begin{aligned}
\frac{\partial \zeta}{\partial t} + f\delta + \beta v &= 0 \\
\frac{\partial \delta}{\partial t} - f\zeta + \beta u + \nabla^2 P &= 0 \\
\frac{\partial P}{\partial t} + gh\delta &= 0\,.
\end{aligned}
$$

This system is equivalent to the Laplace Tidal Equations. No additional approximations have been made . . .

. . . however, the vorticity and divergence forms enable us to examine various simple approximate solutions.

Mathematical Interlude

The eignefunctions of the Laplacian operator on the sphere are called spherical harmonics:

$$Y_n^m(\lambda, \phi) = \exp(im\lambda) P_n^m(\phi)$$

where $P_n^m(\phi)$ are the associated Legendre functions.

Mathematical Interlude

The eignefunctions of the Laplacian operator on the sphere are called spherical harmonics:

$$Y_n^m(\lambda, \phi) = \exp(im\lambda) P_n^m(\phi)$$

where $P_n^m(\phi)$ are the associated Legendre functions.

We have

$$\nabla^2 Y_n^m = -\frac{n(n+1)}{a^2} Y_n^m \, .$$

Mathematical Interlude

The eignefunctions of the Laplacian operator on the sphere are called **spherical harmonics**:

$$Y_n^m(\lambda, \phi) = \exp(im\lambda) P_n^m(\phi)$$

where $P_n^m(\phi)$ are the **associated Legendre functions.**

We have

$$\nabla^2 Y_n^m = -\frac{n(n+1)}{a^2} Y_n^m \, .$$

The **zonal wavenumber** is m. The **total waveno.** is n.

The 'beta-term' in the vorticity equation is

$$\beta v = \frac{2\Omega\cos\phi}{a}\left(\frac{1}{a\cos\phi}\frac{\partial\psi}{\partial\lambda} + \frac{1}{a}\frac{\partial\chi}{\partial\lambda}\right)$$

The ‘beta-term’ in the vorticity equation is

$$\beta v = \frac{2\Omega\cos\phi}{a}\left(\frac{1}{a\cos\phi}\frac{\partial\psi}{\partial\lambda} + \frac{1}{a}\frac{\partial\chi}{\partial\lambda}\right)$$

For quasi-non-divergent flow ($|\delta| \ll |\zeta|$) it becomes

$$\beta v \approx \frac{2\Omega}{a^2}\frac{\partial\psi}{\partial\lambda}$$

Rossby-Haurwitz Modes

If we suppose that the solution is quasi-nondivergent (that is, $|\delta| \ll |\zeta|$), the wind is given approximately in terms of the stream function $(u, v) \approx (-\psi_y, \psi_x)$.

Rossby-Haurwitz Modes

If we suppose that the solution is quasi-nondivergent (that is, $|\delta| \ll |\zeta|$), the wind is given approximately in terms of the stream function $(u, v) \approx (-\psi_y, \psi_x)$.

The vorticity equation becomes

$$\nabla^2\psi_t + \beta\psi_x = O(\delta)\,,$$

and we can ignore the right-hand side.

Rossby-Haurwitz Modes

If we suppose that the solution is quasi-nondivergent (that is, $|\delta| \ll |\zeta|$), the wind is given approximately in terms of the stream function $(u, v) \approx (-\psi_y, \psi_x)$.

The vorticity equation becomes

$$\nabla^2\psi_t + \beta\psi_x = O(\delta) \,,$$

and we can ignore the right-hand side.

Assuming the stream function has the wave-like structure of a spherical harmonic, we substitute

$$\psi = \psi_0 Y_n^m(\lambda, \phi) \exp(-i\nu t)$$

in the vorticity equation, and obtain the frequency:

$$\nu = \nu_R \equiv -\frac{2\Omega m}{n(n+1)} \,.$$

$$\nu = \nu_R \equiv -\frac{2\Omega m}{n(n+1)}\,.$$

This is the celebrated dispersion relation for Rossby-Haurwitz waves (Haurwitz, 1940).

$$\nu = \nu_R \equiv -\frac{2\Omega m}{n(n+1)} .$$

This is the celebrated dispersion relation for **Rossby-Haurwitz waves** (Haurwitz, 1940).

We can ignore sphericity (the β-plane approximation) and assume harmonic dependence

$$\psi(x, y, t) = \psi_0 \exp[i(kx + \ell y - \nu t)] ,$$

$$\nu = \nu_R \equiv -\frac{2\Omega m}{n(n+1)} .$$

This is the celebrated dispersion relation for Rossby-Haurwitz waves (Haurwitz, 1940).

We can ignore sphericity (the β-plane approximation) and assume harmonic dependence

$$\psi(x, y, t) = \psi_0 \exp[i(kx + \ell y - \nu t)] ,$$

Then the dispersion relation is

$$c = \frac{\nu}{k} = -\frac{\beta}{k^2 + \ell^2} ,$$

which is the phase-speed found by Rossby (1939).

The Rossby or Rossby-Haurwitz waves are, to the first approximation, non-divergent waves which travel westward, the phase speed being greatest for the waves of largest scale.

The Rossby or Rossby-Haurwitz waves are, to the first approximation, non-divergent waves which travel westward, the phase speed being greatest for the waves of largest scale.

The RH waves are of relatively low frequency — $|\nu| \leq \Omega$ — and the frequency decreases as the spatial scale decreases.

The Rossby or Rossby-Haurwitz waves are, to the first approximation, non-divergent waves which travel westward, the phase speed being greatest for the waves of largest scale.

The RH waves are of relatively low frequency — $|\nu| \leq \Omega$ — and the frequency decreases as the spatial scale decreases.

We may write the divergence equation as

$$\nabla^2 P - f\zeta - \beta\psi_y = O(\delta) \,.$$

The Rossby or Rossby-Haurwitz waves are, to the first approximation, non-divergent waves which travel westward, the phase speed being greatest for the waves of largest scale.

The RH waves are of relatively low frequency — $|\nu| \leq \Omega$ — and the frequency decreases as the spatial scale decreases.

We may write the divergence equation as

$$\nabla^2 P - f\zeta - \beta\psi_y = O(\delta) .$$

Ignoring the r.h.s., we get the linear balance equation

$$\nabla^2 P = \nabla \cdot f \nabla \psi ,$$

a diagnostic relationship between the geopotential and the stream function.

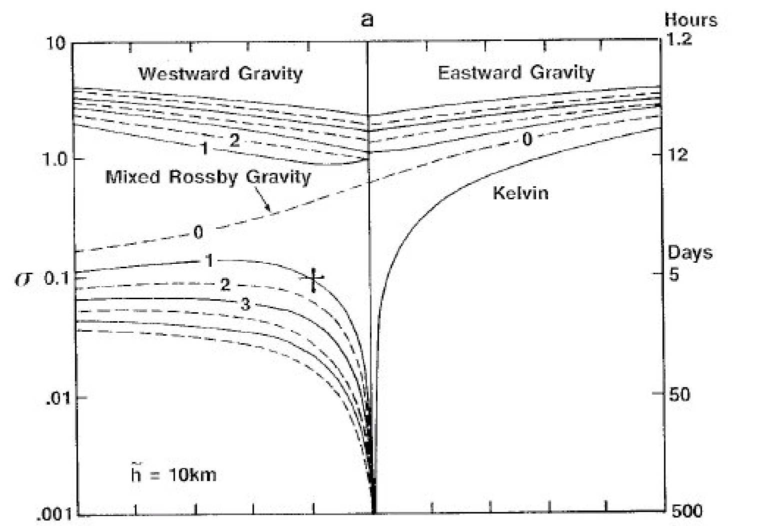
This also follows immediately from the assumption that the wind is both non-divergent and geostrophic:

$$\mathbf{V} = \mathbf{k} \times \nabla\psi \qquad \text{and} \qquad f\mathbf{V} = \mathbf{k} \times \nabla P$$

This also follows immediately from the assumption that the wind is both non-divergent and geostrophic:

$$\mathbf{V} = \mathbf{k} \times \nabla\psi \qquad \text{and} \qquad f\mathbf{V} = \mathbf{k} \times \nabla P$$

If variations of f are ignored, we can assume $P = f\psi$. The wind and pressure are in approximate geostrophic balance for Rossby-Haurwitz waves.



The eigenmodes of the Laplace Tidal Equations ($h = 10$ km).

Gravity Wave Modes

We assume now that the solution is **quasi-irrotational**, *i.e.* that $|\zeta| \ll |\delta|$.

Gravity Wave Modes

We assume now that the solution is quasi-irrotational, *i.e.* that $|\zeta| \ll |\delta|$.

Then the wind is given approximately by $(u, v) \approx (\chi_x, \chi_y)$ and the divergence equation becomes

$$\nabla^2 \chi_t + \beta \chi_x + \nabla^2 P = O(\zeta)$$

with the right-hand side negligible.

Gravity Wave Modes

We assume now that the solution is quasi-irrotational, *i.e.* that $|\zeta| \ll |\delta|$.

Then the wind is given approximately by $(u, v) \approx (\chi_x, \chi_y)$ and the divergence equation becomes

$$\nabla^2 \chi_t + \beta \chi_x + \nabla^2 P = O(\zeta)$$

with the right-hand side negligible.

Using the continuity equation to eliminate P, we get

$$\nabla^2 \chi_{tt} + \beta \chi_{xt} - gh \nabla^4 \chi = 0\,.$$

Gravity Wave Modes

We assume now that the solution is **quasi-irrotational**, *i.e.* that $|\zeta| \ll |\delta|$.

Then the wind is given approximately by $(u, v) \approx (\chi_x, \chi_y)$ and the divergence equation becomes

$$\nabla^2 \chi_t + \beta \chi_x + \nabla^2 P = O(\zeta)$$

with the right-hand side negligible.

Using the continuity equation to eliminate P, we get

$$\nabla^2 \chi_{tt} + \beta \chi_{xt} - gh \nabla^4 \chi = 0\,.$$

If we look for a solution $\chi = \chi_0 Y_n^m(\lambda, \phi) \exp(-i\nu t)$ we find that

$$\nu^2 + \left(-\frac{2\Omega m}{n(n+1)} \right) \nu - \frac{n(n+1)gh}{a^2} = 0\,.$$

$$\nu^2 + \left(-\frac{2\Omega m}{n(n+1)}\right)\nu - \frac{n(n+1)gh}{a^2} = 0\,.$$

$$\nu^2 + \left(-\frac{2\Omega m}{n(n+1)}\right)\nu - \frac{n(n+1)gh}{a^2} = 0\,.$$

The coefficient of the second term is just the Rossby-Haurwitz frequency ν_R, so that

$$\nu = \pm\sqrt{\nu_G^2 + (\tfrac{1}{2}\nu_R)^2} - \tfrac{1}{2}\nu_R\,, \qquad \text{where} \qquad \nu_G \equiv \sqrt{\frac{n(n+1)gh}{a^2}}\,,$$

$$\nu^2 + \left(-\frac{2\Omega m}{n(n+1)}\right)\nu - \frac{n(n+1)gh}{a^2} = 0\,.$$

The coefficient of the second term is just the Rossby-Haurwitz frequency ν_R, so that

$$\nu = \pm\sqrt{\nu_G^2 + (\tfrac{1}{2}\nu_R)^2} - \tfrac{1}{2}\nu_R\,, \qquad \text{where} \qquad \nu_G \equiv \sqrt{\frac{n(n+1)gh}{a^2}}\,,$$

Noting that $|\nu_G| \gg |\nu_R|$, it follows that

$$\nu_\pm \approx \pm\nu_G = \pm\sqrt{\frac{n(n+1)gh}{a^2}}\,,$$

the frequency of pure gravity waves.

$$\nu^2 + \left(-\frac{2\Omega m}{n(n+1)}\right)\nu - \frac{n(n+1)gh}{a^2} = 0\,.$$

The coefficient of the second term is just the Rossby-Haurwitz frequency ν_R, so that

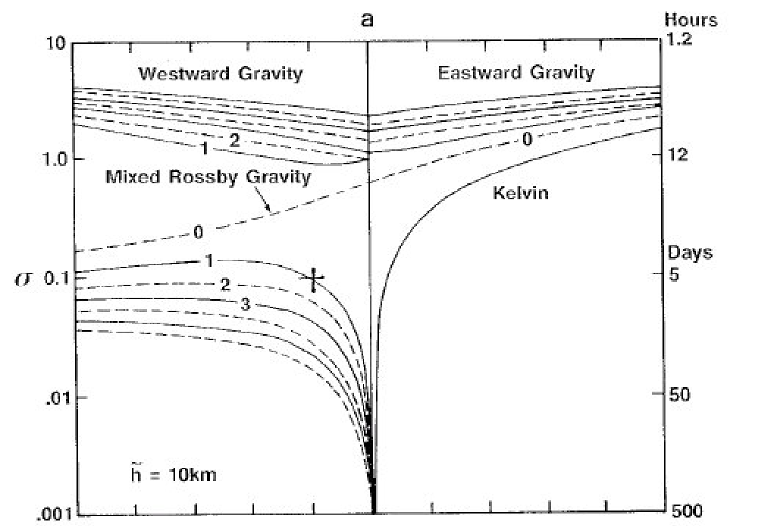
$$\nu = \pm\sqrt{\nu_G^2 + (\tfrac{1}{2}\nu_R)^2} - \tfrac{1}{2}\nu_R\,, \qquad \text{where} \qquad \nu_G \equiv \sqrt{\frac{n(n+1)gh}{a^2}}\,,$$

Noting that $|\nu_G| \gg |\nu_R|$, it follows that

$$\nu_\pm \approx \pm\nu_G = \pm\sqrt{\frac{n(n+1)gh}{a^2}}\,,$$

the frequency of pure gravity waves.

There are then two solutions, representing waves travelling eastward and westward with equal speeds. The frequency increases approximately linearly with the total wavenumber n.



The eigenmodes of the Laplace TIdal Equations ($h = 10\,$km).

Outline

Reminder on linear algebra

Let **L** be a matrix. An eigenvector **e** of **L** with eigenvalue λ satisfies

$$\mathbf{L}\mathbf{e} = \lambda\mathbf{e}$$

In general there are n eigenvectors for an $n \times n$ matrix.

Reminder on linear algebra

Let $\mathbf{L}$ be a matrix. An **eigenvector** $\mathbf{e}$ of $\mathbf{L}$ with eigenvalue λ satisfies

$$\mathbf{L}\mathbf{e} = \lambda\mathbf{e}$$

In general there are n eigenvectors for an $n \times n$ matrix.

We form the eigenvector and eigenvalue matrices

$$\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] \qquad \text{and} \qquad \Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$$

Then the eigenvector relationships can be written as

$$\mathbf{L}\mathbf{E} = \mathbf{E}\Lambda$$

Then the eigenvector relationships can be written as

$$\mathbf{L}\mathbf{E} = \mathbf{E}\Lambda$$

For a symmetric matrix, the eigenvalues are real and the eigenvectors are orthogonal:

$$\mathbf{E}\mathbf{E}^{\mathrm{T}} = \mathbf{E}^{\mathrm{T}}\mathbf{E} = \mathbf{I}\,.$$

It follows immediately that

$$\mathbf{L} = \mathbf{E}\Lambda\mathbf{E}^{\mathrm{T}} \qquad \text{and} \qquad \mathbf{E}^{\mathrm{T}}\mathbf{L}\mathbf{E} = \Lambda\,.$$

Normal Mode Initialization ⋆

Let $\mathbf{X}(t)$ be the state vector of dependent variables. The model equations can be written schematically as

$$\dot{\mathbf{X}} + i\mathbf{L}\mathbf{X} + \mathbf{N}(\mathbf{X}) = \mathbf{0}$$

with $\mathbf{L}$ a matrix and $\mathbf{N}$ a nonlinear vector function.

Normal Mode Initialization ⋆

Let $\mathbf{X}(t)$ be the state vector of dependent variables. The model equations can be written schematically as

$$\dot{\mathbf{X}} + i\mathbf{L}\mathbf{X} + \mathbf{N}(\mathbf{X}) = \mathbf{0}$$

with **L** a matrix and **N** a nonlinear vector function.

Denote the eigenvector matrix of **L** by **E** and the diagonal eigenvalue matrix as Λ. Then

$$\mathbf{E}^{\mathrm{T}}\mathbf{L}\mathbf{E} = \Lambda \,.$$

We introduce a transformed state vector

$$\mathbf{W} = \mathbf{E}^{\mathrm{T}}\mathbf{X}$$

and multiply the model equations on the left by $\mathbf{E}^{\mathrm{T}}$.

$$\mathbf{E}^{\mathrm{T}}\dot{\mathbf{X}} + i\mathbf{E}^{\mathrm{T}}\mathbf{L}\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{X} + \mathbf{E}^{\mathrm{T}}\mathbf{N}(\mathbf{X}) = \mathbf{0}$$

We introduce a transformed state vector

$$\mathbf{W} = \mathbf{E}^{\mathrm{T}}\mathbf{X}$$

and multiply the model equations on the left by $\mathbf{E}^{\mathrm{T}}$.

$$\mathbf{E}^{\mathrm{T}}\dot{\mathbf{X}} + i\mathbf{E}^{\mathrm{T}}\mathbf{L}\mathbf{E}\mathbf{E}^{\mathrm{T}}\mathbf{X} + \mathbf{E}^{\mathrm{T}}\mathbf{N}(\mathbf{X}) = \mathbf{0}$$

This may be written

$$\dot{\mathbf{W}} + i\Lambda\mathbf{W} + \hat{\mathbf{N}}(\mathbf{X}) = \mathbf{0}$$

where $\hat{\mathbf{N}} = \mathbf{E}^{\mathrm{T}}\mathbf{N}(\mathbf{X})$. Recall that Λ is diagonal.

This linear system seperates into two subsystems.

The eigenvalues fall in to **slow** and **fast** subsets. We partition the eigenvalue matrix on this basis:

$$\Lambda = \begin{bmatrix} \Lambda_Y & \mathbf{0} \\ \mathbf{0} & \Lambda_Z \end{bmatrix}$$

where Λ_Y and Λ_Z are diagonal matrices of eigenfrequencies for the two types of modes.

The eigenvalues fall in to **slow** and **fast** subsets.
We partition the eigenvalue matrix on this basis:

$$\Lambda = \begin{bmatrix} \Lambda_Y & 0 \\ 0 & \Lambda_Z \end{bmatrix}$$

where Λ_Y and Λ_Z are diagonal matrices of eigenfrequencies for the two types of modes.

The state vector W is comprised of two sub-vectors:

$$W = \begin{bmatrix} Y \\ Z \end{bmatrix}$$

The eigenvalues fall in to slow and fast subsets. We partition the eigenvalue matrix on this basis:

$$\Lambda = \begin{bmatrix} \Lambda_Y & 0 \\ 0 & \Lambda_Z \end{bmatrix}$$

where Λ_Y and Λ_Z are diagonal matrices of eigenfrequencies for the two types of modes.

The state vector W is comprised of two sub-vectors:

$$W = \begin{bmatrix} Y \\ Z \end{bmatrix}$$

The system then separates into two subsystems, for the low and high frequency components.

$$\begin{aligned} \dot{\mathbf{Y}} + i\Lambda_{\mathbf{Y}}\mathbf{Y} + \mathbf{N}_Y(\mathbf{Y},\mathbf{Z}) &= 0 \\ \dot{\mathbf{Z}} + i\Lambda_{\mathbf{Z}}\mathbf{Z} + \mathbf{N}_Z(\mathbf{Y},\mathbf{Z}) &= 0 \end{aligned}$$

The vectors Y and Z are the coefficients of the LF and HF components of the flow: the slow and fast components.

$$
\begin{aligned}
\dot{\mathbf{Y}} + i\Lambda_{\mathbf{Y}}\mathbf{Y} + \mathbf{N}_Y(\mathbf{Y},\mathbf{Z}) &= \mathbf{0} \\
\dot{\mathbf{Z}} + i\Lambda_{\mathbf{Z}}\mathbf{Z} + \mathbf{N}_Z(\mathbf{Y},\mathbf{Z}) &= \mathbf{0}
\end{aligned}
$$

The vectors Y and Z are the coefficients of the LF and HF components of the flow: the slow and fast components.

Let us now suppose that the initial fields are separated into slow and fast parts.

The fast modes may be removed so as to leave only the Rossby waves:

Replace $\mathbf{W} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{Z} \end{bmatrix}$ by $\mathbf{W} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{0} \end{bmatrix}$ at time $t = 0$.

It might be hoped that this process of **linear normal mode initialization**, imposing the condition

$$\text{LNMI:} \qquad \mathbf{Z} = \mathbf{0} \quad \text{at} \quad t = 0$$

would ensure a noise-free forecast.

However, the results are disappointing: the noise is reduced initially, but soon reappears.

It might be hoped that this process of linear normal mode initialization, imposing the condition

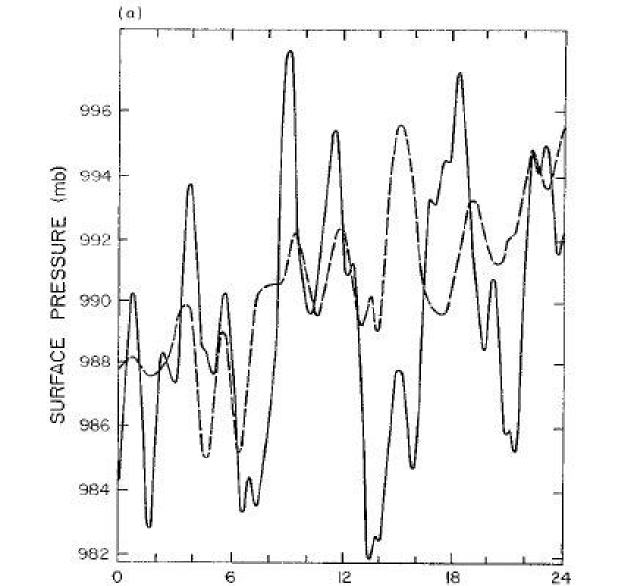
$$\textbf{LNMI:} \qquad \mathbf{Z} = \mathbf{0} \quad \textbf{at} \quad t = 0$$

would ensure a noise-free forecast.

However, the results are disappointing: the noise is reduced initially, but soon reappears.

The equations are nonlinear, and the slow components interact nonlinearly in such a way as to generate gravity waves.

The problem of noise remains: the gravity waves are small to begin with, but they grow rapidly.



Surface pressure evolution: No Initialization and LNMI.

To control the growth of HF components, Bennert Machenhauer (1977) proposed setting their initial rate-of-change to zero, in the hope that they would remain small throughout the forecast.

Ferd Baer (1977) proposed a somewhat more general method, using a two-timing perturbation technique.

To control the growth of HF components, Bennert Machenhauer (1977) proposed setting their initial rate-of-change to zero, in the hope that they would remain small throughout the forecast.

Ferd Baer (1977) proposed a somewhat more general method, using a two-timing perturbation technique.

The forecast, starting from initial fields modified so that $\dot{\mathbf{Z}} = \mathbf{0}$ at $t = 0$ is very smooth and the spurious gravity wave oscillations are almost completely removed.

NNMI: $\dot{\mathbf{Z}} = \mathbf{0}$ **at** $t = 0$

Applying NNMI to the the equation for the fast modes:

$$\dot{\mathbf{Z}} + i\Lambda_{\mathbf{Z}}\mathbf{Z} + \hat{\mathbf{N}}_Z(\mathbf{Y},\mathbf{Z}) = \mathbf{0}$$

we get

$$i\Lambda_{\mathbf{Z}}\mathbf{Z} + \hat{\mathbf{N}}_Z(\mathbf{Y},\mathbf{Z}) = \mathbf{0} \qquad \text{or} \qquad \mathbf{Z} = i\Lambda_{\mathbf{Z}}^{-1}\hat{\mathbf{N}}_Z(\mathbf{Y},\mathbf{Z})$$

NNMI: $\dot{\mathbf{Z}} = \mathbf{0}$ **at** $t = 0$

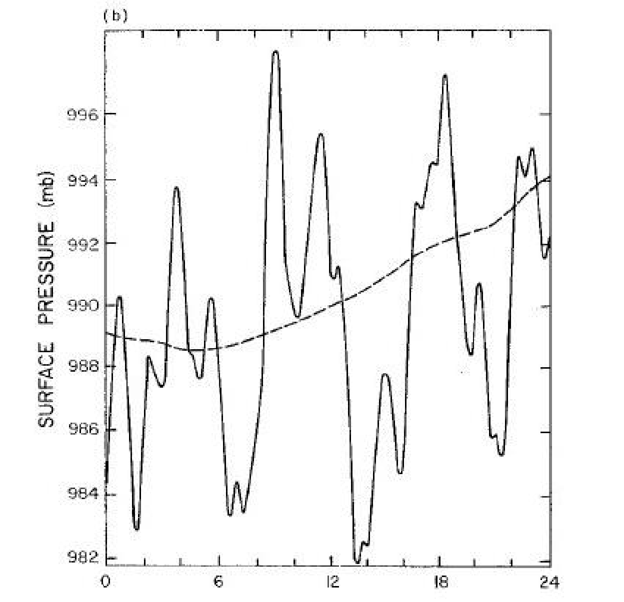
Applying NNMI to the the equation for the fast modes:

$$\dot{\mathbf{Z}} + i\Lambda_{\mathbf{Z}}\mathbf{Z} + \hat{\mathbf{N}}_Z(\mathbf{Y}, \mathbf{Z}) = \mathbf{0}$$

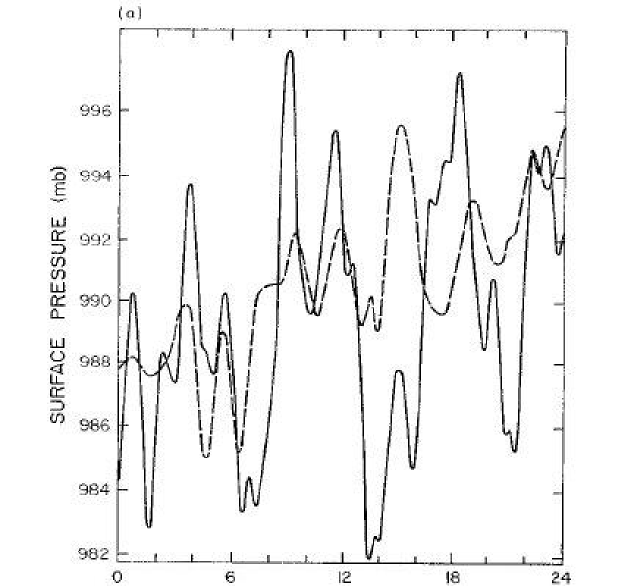
we get

$$i\Lambda_{\mathbf{Z}}\mathbf{Z} + \hat{\mathbf{N}}_Z(\mathbf{Y}, \mathbf{Z}) = \mathbf{0} \qquad \text{or} \qquad \mathbf{Z} = i\Lambda_{\mathbf{Z}}^{-1}\hat{\mathbf{N}}_Z(\mathbf{Y}, \mathbf{Z})$$

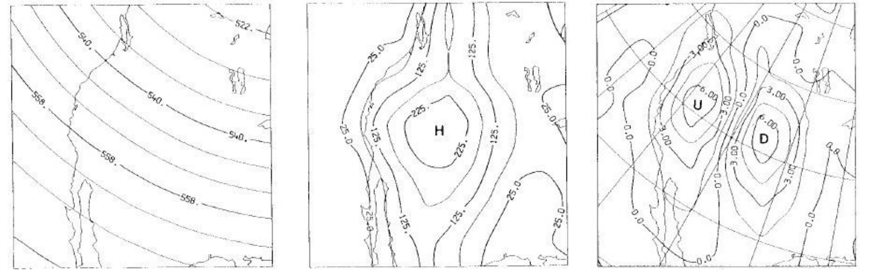
The method takes account of the nonlinear nature of the equations, and is referred to as nonlinear normal mode initialization.



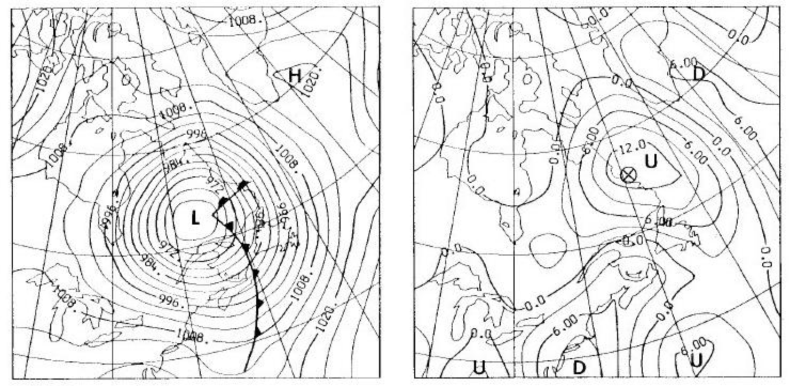
Surface pressure evolution: No Initialization and NNMI.



Surface pressure evolution: No Initialization and LNMI.



Vertical velocity *w* for flow over the Rockies.
A realistic *w* field is generated by nonlinear normal mode initialization.



Generation of vertical velocity *w* in frontal depression.
A realistic *w* field is generated by nonlinear normal mode initialization.

Outline

Example: The Swinging Spring

We will illustrate linear normal mode initialization (LNMI) and nonlinear normal mode initialization (NNMI) by application to a simple mechanical system.

Example: The Swinging Spring

We will illustrate linear normal mode initialization (LNMI) and nonlinear normal mode initialization (NNMI) by application to a simple mechanical system.

The *swinging spring* comprises a heavy bob suspended by a light elastic spring. The bob is free to move in a vertical plane.

Example: The Swinging Spring

We will illustrate linear normal mode initialization (LNMI) and nonlinear normal mode initialization (NNMI) by application to a simple mechanical system.

The *swinging spring* comprises a heavy bob suspended by a light elastic spring. The bob is free to move in a vertical plane.

The oscillations of this system are of two types, distinguished by their restoring mechanisms.

Example: The Swinging Spring

We will illustrate linear normal mode initialization (LNMI) and nonlinear normal mode initialization (NNMI) by application to a simple mechanical system.

The *swinging spring* comprises a heavy bob suspended by a light elastic spring. The bob is free to move in a vertical plane.

The oscillations of this system are of two types, distinguished by their restoring mechanisms.

We consider the elastic oscillations to be analogues of the HF gravity waves in the atmosphere.

Example: The Swinging Spring

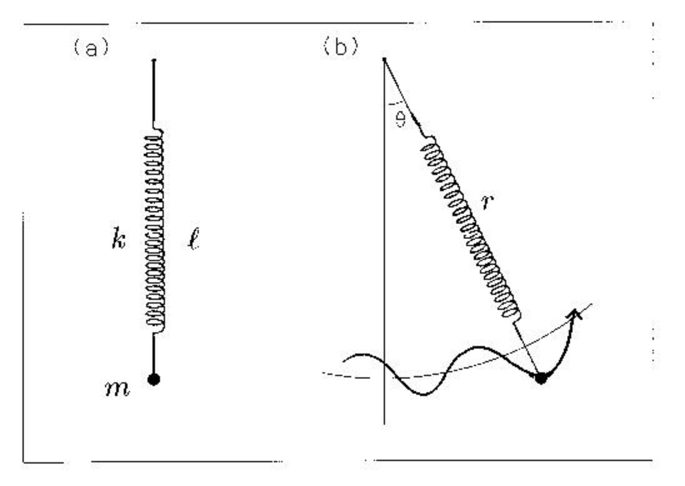
We will illustrate linear normal mode initialization (LNMI) and nonlinear normal mode initialization (NNMI) by application to a simple mechanical system.

The *swinging spring* comprises a heavy bob suspended by a light elastic spring. The bob is free to move in a vertical plane.

The oscillations of this system are of two types, distinguished by their restoring mechanisms.

We consider the elastic oscillations to be analogues of the HF gravity waves in the atmosphere.

Similarly, the LF rotational motions correspond to the rotational or Rossby waves.



The swinging spring (2D case)

The Dynamical Equations

Let ℓ_0 be the unstretched length of the spring, k its elasticity or stiffness and m the mass of the bob.

The Dynamical Equations

Let ℓ_0 be the unstretched length of the spring, k its elasticity or stiffness and m the mass of the bob.

At equilibrium the balance of forces is:

$$k(\ell - \ell_0) = mg$$

The Dynamical Equations

Let ℓ_0 be the unstretched length of the spring, k its elasticity or stiffness and m the mass of the bob.

At equilibrium the balance of forces is:

$$k(\ell - \ell_0) = mg$$

Polar coordinates $q_r = r$ and $q_\theta = \theta$ are used, and the radial and angular momenta are $p_r = m\dot{r}$, $p_\theta = mr^2\dot{\theta}$.

The Dynamical Equations

Let ℓ_0 be the unstretched length of the spring, k its elasticity or stiffness and m the mass of the bob.

At equilibrium the balance of forces is:

$$k(\ell - \ell_0) = mg$$

Polar coordinates $q_r = r$ and $q_\theta = \theta$ are used, and the radial and angular momenta are $p_r = m\dot{r}$, $p_\theta = mr^2\dot{\theta}$.

The Hamiltonian is (in this case) the sum of kinetic, elastic potential and gravitational potential energy:

$$H = \frac{1}{2m}\left(p_r^2 + \frac{p_\theta^2}{r^2}\right) + \tfrac{1}{2}k(r-\ell_0)^2 - mgr\cos\theta\,.$$

The (canonical) dynamical equations are

$$
\begin{aligned}
\dot\theta &= p_\theta / mr^2 \\
\dot p_\theta &= -mgr\sin\theta \\
\dot r &= p_r / m \\
\dot p_r &= p_\theta^2 / mr^3 - k(r-\ell_0) + mg\cos\theta \,.
\end{aligned}
$$

The (canonical) dynamical equations are

$$\begin{aligned}
\dot{\theta} &= p_\theta / mr^2 \\
\dot{p}_\theta &= -mgr\sin\theta \\
\dot{r} &= p_r / m \\
\dot{p}_r &= p_\theta^2 / mr^3 - k(r - \ell_0) + mg\cos\theta\,.
\end{aligned}$$

These equations may be written symbolically as

$$\dot{\mathbf{X}} + \mathbf{L}\mathbf{X} + \mathbf{N}(\mathbf{X}) = 0$$

where $\mathbf{X} = (\theta, p_\theta, r, p_r)^{\mathrm{T}}$, **L** is the matrix for the linear terms and **N** is a nonlinear vector function.

The (canonical) dynamical equations are

$$\begin{aligned} \dot{\theta} &= p_\theta / mr^2 \\ \dot{p}_\theta &= -mgr \sin\theta \\ \dot{r} &= p_r / m \\ \dot{p}_r &= p_\theta^2 / mr^3 - k(r - \ell_0) + mg\cos\theta \,. \end{aligned}$$

These equations may be written symbolically as

$$\dot{\mathbf{X}} + \mathbf{L}\mathbf{X} + \mathbf{N}(\mathbf{X}) = 0$$

where $\mathbf{X} = (\theta, p_\theta, r, p_r)^{\mathrm{T}}$, **L** is the matrix for the linear terms and **N** is a nonlinear vector function.

We now suppose that the amplitude of the motion is small, so that $|r'| = |r - \ell| \ll \ell$ and $|\theta| \ll 1$.

The (canonical) dynamical equations are

$$
\begin{aligned}
\dot{\theta} &= p_\theta / mr^2 \\
\dot{p}_\theta &= -mgr\sin\theta \\
\dot{r} &= p_r / m \\
\dot{p}_r &= p_\theta^2 / mr^3 - k(r-\ell_0) + mg\cos\theta\,.
\end{aligned}
$$

These equations may be written symbolically as

$$\dot{\mathbf{X}} + \mathbf{L}\mathbf{X} + \mathbf{N}(\mathbf{X}) = 0$$

where $\mathbf{X} = (\theta, p_\theta, r, p_r)^{\mathrm{T}}$, L is the matrix for the linear terms and N is a nonlinear vector function.

We now suppose that the amplitude of the motion is small, so that $|r'| = |r - \ell| \ll \ell$ and $|\theta| \ll 1$.

The state vector X comprises two sub-vectors:

$$\mathbf{X} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}, \qquad \text{where} \quad \mathbf{Y} = \begin{pmatrix} \theta \\ p_\theta \end{pmatrix} \quad \text{and} \quad \mathbf{Z} = \begin{pmatrix} r' \\ p_r \end{pmatrix}.$$

We call the motion described by Y the rotational component and that described by Z the elastic component.

We call the motion described by Y the rotational component and that described by Z the elastic component.

The rotational equations may be written

$$\ddot{\theta} + (g/\ell)\theta = 0$$

which is the equation for a simple pendulum having oscillatory solutions with frequency

$$\omega_{\mathrm{R}} \equiv \sqrt{\frac{g}{\ell}} \, .$$

We call the motion described by Y the rotational component and that described by Z the elastic component.

The rotational equations may be written

$$\ddot{\theta} + (g/\ell)\theta = 0$$

which is the equation for a simple pendulum having oscillatory solutions with frequency

$$\omega_{\mathrm{R}} \equiv \sqrt{\frac{g}{\ell}} \,.$$

The remaining two equations yield

$$\ddot{r}' + (k/m)r' = 0 \,,$$

the equations for elastic oscillations with frequency

$$\omega_{\mathrm{E}} = \sqrt{\frac{k}{m}} \,.$$

We define the ratio of the rotational and elastic frequencies:

$$\omega_{\mathrm{R}} = \sqrt{\frac{g}{\ell}}\,, \qquad \omega_{\mathrm{E}} = \sqrt{\frac{k}{m}}\,, \qquad \epsilon \equiv \left(\frac{\omega_{\mathrm{R}}}{\omega_{\mathrm{E}}}\right)\,.$$

We define the **ratio of the rotational and elastic frequencies:**

$$\omega_{\rm R} = \sqrt{\frac{g}{\ell}}\,, \qquad \omega_{\rm E} = \sqrt{\frac{k}{m}}\,, \qquad \epsilon \equiv \left(\frac{\omega_{\rm R}}{\omega_{\rm E}}\right)\,.$$

It is easily shown that $\epsilon < 1$, so the rotational frequency is always less than the elastic.

$$\epsilon^2 = \frac{mg}{k\ell} = \left(1 - \frac{\ell_0}{\ell}\right) < 1\,, \qquad \text{so that} \qquad |\omega_{\rm R}| < |\omega_{\rm E}|\,.$$

We define the ratio of the rotational and elastic frequencies:

$$\omega_{\mathrm{R}} = \sqrt{\frac{g}{\ell}}\,, \qquad \omega_{\mathrm{E}} = \sqrt{\frac{k}{m}}\,, \qquad \epsilon \equiv \left(\frac{\omega_{\mathrm{R}}}{\omega_{\mathrm{E}}}\right)\,.$$

It is easily shown that $\epsilon < 1$, so the rotational frequency is always less than the elastic.

$$\epsilon^2 = \frac{mg}{k\ell} = \left(1 - \frac{\ell_0}{\ell}\right) < 1\,, \qquad \text{so that} \qquad |\omega_{\mathrm{R}}| < |\omega_{\mathrm{E}}|\,.$$

We assume that the parameters are such that

$$\epsilon \ll 1$$

Then the linear normal modes are clearly distinct:

- The rotational mode has low frequency (LF)
- The elastic mode has high frequency (HF).

Linear and Nonlinear Initialization

For small amplitude motions the LF and HF oscillations are independent of each other. They evolve without interaction.

Linear and Nonlinear Initialization

For small amplitude motions the LF and HF oscillations are independent of each other. They evolve without interaction.

We can suppress the HF component completely by setting its initial amplitude to zero:

$$\mathbf{Z} = (r', p_r)^{\mathrm{T}} = \mathbf{0} \quad \textbf{at} \quad t = 0\,.$$

This procedure is called linear initialization.

Linear and Nonlinear Initialization

For small amplitude motions the LF and HF oscillations are independent of each other. They evolve without interaction.

We can suppress the HF component completely by setting its initial amplitude to zero:

$$\mathbf{Z} = (r', p_r)^{\mathrm{T}} = \mathbf{0} \quad \textbf{at} \quad t = 0\,.$$

This procedure is called linear initialization.

When the amplitude is large, nonlinear terms are no longer negligible. The LF and HF motions interact.

Linear and Nonlinear Initialization

For small amplitude motions the LF and HF oscillations are independent of each other. They evolve without interaction.

We can suppress the HF component completely by setting its initial amplitude to zero:

$$\mathbf{Z} = (r', p_r)^{\mathrm{T}} = \mathbf{0} \quad \textbf{at} \quad t = 0\,.$$

This procedure is called linear initialization.

When the amplitude is large, nonlinear terms are no longer negligible. The LF and HF motions interact.

It is clear from the equations that linear initialization will not ensure permanent absence of HF motions . . .
. . . the nonlinear LF terms generate radial momentum.

To achieve better results, we set the initial tendency of the HF components to zero:

$$\dot{\mathbf{Z}} = (\dot{r}, \dot{p}_r)^{\mathrm{T}} = \mathbf{0} \quad \textbf{at} \quad t = 0\,,$$

This procedure is called nonlinear initialization.

To achieve better results, we set the **initial tendency** of the HF components to zero:

$$\dot{\mathbf{Z}} = (\dot{r}, \dot{p}_r)^{\mathrm{T}} = \mathbf{0} \quad \textbf{at} \quad t = 0\,,$$

This procedure is called **nonlinear initialization.**

For the spring, we can deduce **explicit expressions** for the initial conditions:

$$r(0) = r_{\mathrm{B}} \equiv \frac{\ell(1 - \epsilon^2(1 - \cos\theta))}{1 - (\dot{\theta}/\omega_{\mathrm{E}})^2}\,, \qquad p_r(0) = 0\,.$$

To achieve better results, we set the **initial tendency** of the HF components to zero:

$$\dot{\mathbf{Z}} = (\dot{r}, \dot{p}_r)^{\mathrm{T}} = \mathbf{0} \quad \textbf{at} \quad t = 0\,,$$

This procedure is called **nonlinear initialization.**

For the spring, we can deduce **explicit expressions** for the initial conditions:

$$r(0) = r_{\mathrm{B}} \equiv \frac{\ell(1-\epsilon^2(1-\cos\theta))}{1-(\dot{\theta}/\omega_{\mathrm{E}})^2}\,, \qquad p_r(0) = 0\,.$$

Thus, given initial conditions $\mathbf{X} = (\theta, p_\theta, r, p_r)^{\mathrm{T}}$, we replace $\mathbf{Z} = (r, p_r)^{\mathrm{T}}$ by $\mathbf{Z}_{\mathrm{B}} = (r_{\mathrm{B}}, 0)^{\mathrm{T}}$.

The rotational component $\mathbf{Y} = (\theta, p_\theta)^{\mathrm{T}}$ is unchanged.

To achieve better results, we set the initial tendency of the HF components to zero:

$$\dot{\mathbf{Z}} = (\dot{r}, \dot{p}_r)^{\mathrm{T}} = \mathbf{0} \quad \textbf{at} \quad t = 0\,,$$

This procedure is called nonlinear initialization.

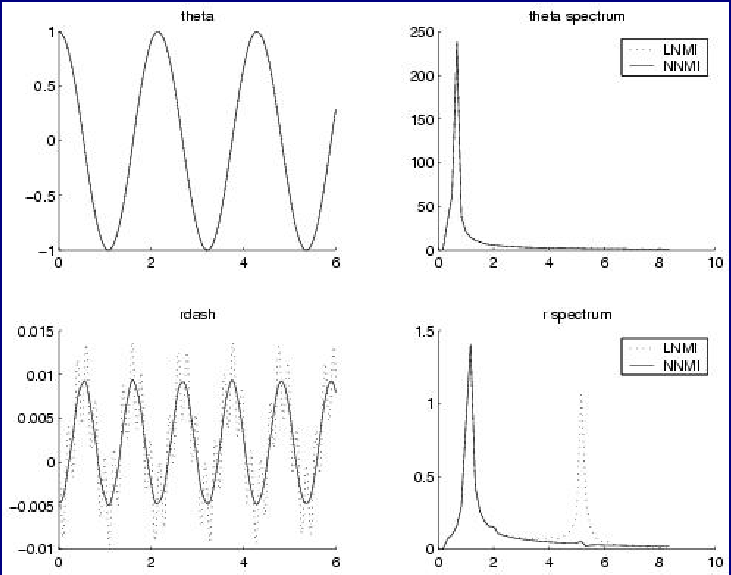
For the spring, we can deduce explicit expressions for the initial conditions:

$$r(0) = r_{\mathrm{B}} \equiv \frac{\ell(1 - \epsilon^2(1 - \cos\theta))}{1 - (\dot{\theta}/\omega_{\mathrm{E}})^2}\,, \qquad p_r(0) = 0\,.$$

Thus, given initial conditions $\mathbf{X} = (\theta, p_\theta, r, p_r)^{\mathrm{T}}$, we replace $\mathbf{Z} = (r, p_r)^{\mathrm{T}}$ by $\mathbf{Z}_{\mathrm{B}} = (r_{\mathrm{B}}, 0)^{\mathrm{T}}$.

The rotational component $\mathbf{Y} = (\theta, p_\theta)^{\mathrm{T}}$ is unchanged.

Does it work? An example shows that it does!



Solution of swinging spring equations for linear (LNMI) and nonlinear (NNMI) initialization.

A Numerical Example

In the accompanying figure, we show the results of two integrations of the spring equations.

A Numerical Example

In the accompanying figure, we show the results of **two integrations** of the spring equations.

The parameter values are

$$m = 1\,, \quad g = \pi^2\,, \quad k = 100\pi^2 \quad \text{and} \quad \ell = 1 \quad \text{(SI units)}$$

Thus, $\epsilon = 0.1$ and the periods of the swinging and springing motions are respectively

$$\tau_{\mathrm{R}} = 2\,\mathrm{s} \qquad \text{and} \qquad \tau_{\mathrm{E}} = 0.2\,\mathrm{s}\,.$$

A Numerical Example

In the accompanying figure, we show the results of **two integrations** of the spring equations.

The parameter values are

$$m = 1\,, \quad g = \pi^2\,, \quad k = 100\pi^2 \quad \text{and} \quad \ell = 1 \quad \text{(SI units)}$$

Thus, $\epsilon = 0.1$ and the periods of the swinging and springing motions are respectively

$$\tau_{\mathrm{R}} = 2\,\mathrm{s} \qquad \text{and} \qquad \tau_{\mathrm{E}} = 0.2\,\mathrm{s}\,.$$

The initial conditions are vanishing velocity ($\dot{r} = \dot{\theta} = 0$), with $\theta(0) = 1$ and $r(0) \in \{1, 0.99540\}$.

A Numerical Example

In the accompanying figure, we show the results of **two integrations** of the spring equations.

The parameter values are

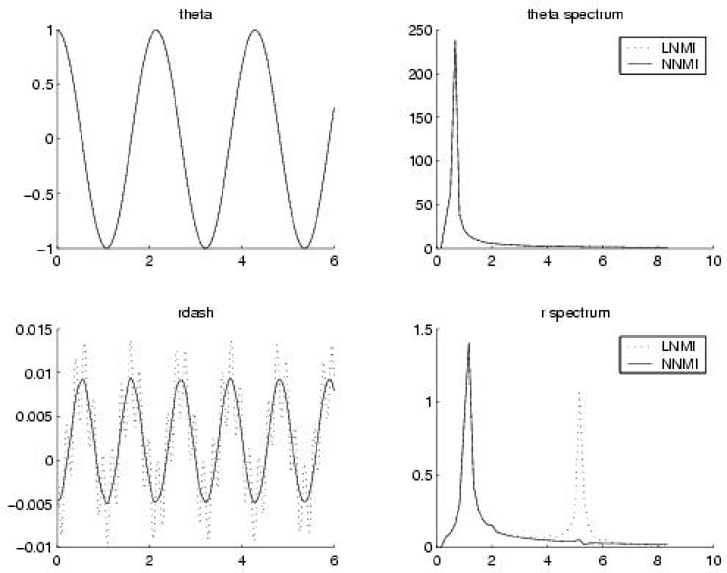
$$m = 1\,, \quad g = \pi^2\,, \quad k = 100\pi^2 \quad \text{and} \quad \ell = 1 \quad \text{(SI units)}$$

Thus, $\epsilon = 0.1$ and the periods of the swinging and springing motions are respectively

$$\tau_{\mathrm{R}} = 2\,\mathrm{s} \qquad \text{and} \qquad \tau_{\mathrm{E}} = 0.2\,\mathrm{s}\,.$$

The initial conditions are vanishing velocity ($\dot{r} = \dot{\theta} = 0$), with $\theta(0) = 1$ and $r(0) \in \{1, 0.99540\}$.

The equations are integrated over a period of 6 s.



Solution of swinging spring equations for linear (LNMI) and nonlinear (NNMI) initialization.

The upper panels show the evolution and spectrum of the slow variable θ. The lower panels are for the fast variable r.

The upper panels show the evolution and spectrum of the slow variable θ. The lower panels are for the fast variable r.

Dotted curves are for linear initialization and solid curves for nonlinear initialization.

The upper panels show the evolution and spectrum of the slow variable θ. The lower panels are for the fast variable r.

Dotted curves are for linear initialization and solid curves for nonlinear initialization.

For slow variable, the curves are indistinguishable.

The spectrum has a clear peak at a frequency of 0.5 cycles per second (Hz).

The upper panels show the evolution and spectrum of the slow variable θ. The lower panels are for the fast variable r.

Dotted curves are for linear initialization and solid curves for nonlinear initialization.

For slow variable, the curves are indistinguishable.

The spectrum has a clear peak at a frequency of 0.5 cycles per second (Hz).

For the fast variable, the linearly initialized evolution has HF noise (dotted curve, lower left panel).

This is confirmed in the spectrum: there is a sharp peak at 5 Hz.

The upper panels show the evolution and spectrum of the slow variable θ. The lower panels are for the fast variable r.

Dotted curves are for linear initialization and solid curves for nonlinear initialization.

For slow variable, the curves are indistinguishable.

The spectrum has a clear peak at a frequency of 0.5 cycles per second (Hz).

For the fast variable, the linearly initialized evolution has HF noise (dotted curve, lower left panel).

This is confirmed in the spectrum: there is a sharp peak at 5 Hz.

When nonlinearly initialized, this peak is removed: only the peak at 1 Hz remains (balanced fast motion).

The balanced fast motion can be understood physically . . .

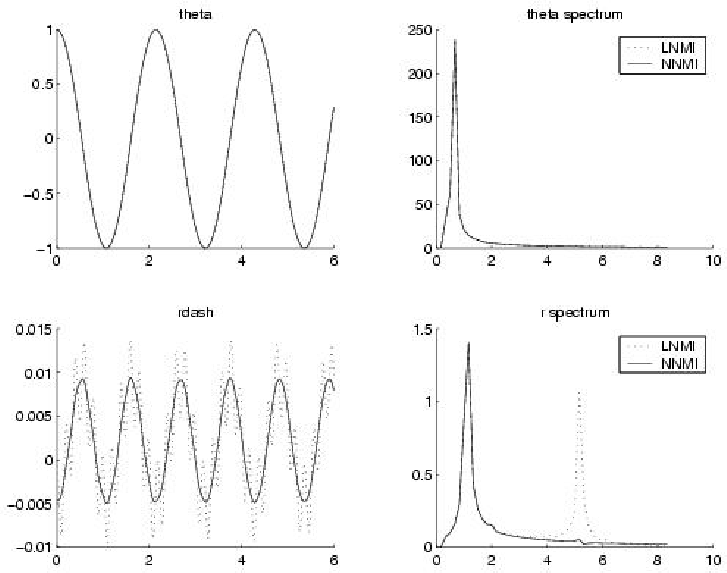
. . . The centrifugal effect stretches the spring twice for each pendular swing: the result is a component of r with a period of one second.

The balanced fast motion can be understood physically . . .

. . . The centrifugal effect stretches the spring twice for each pendular swing: the result is a component of r with a period of one second.

The radial variation does not disappear for balanced motion, but it is of low frequency.

The balanced fast motion is said to be ‘slaved’ (or, better, enslaved) to the slow motion.



Solution of swinging spring equations for linear (LNMI) and nonlinear (NNMI) initialization.

Outline

Introduction to Initialization

Richardson's Forecast

Scale Analysis of the SWE [Skip]

Early Initialization Methods

Laplace Tidal Equations [Skip]

Normal Mode Initialization

The Swinging Spring [Skip]

Digital Filter Initialization

The Notion of Filtering

The concept of filtering has a rôle in many fields, *“from topology to theology, seismology to sociology.”*

The process of filtering involves the selection of those components of an assemblage having some particular property, and the removal or elimination of those that lack it.

A filter is any device or contrivance designed to carry out such a selection.

System Diagram

The input has desired and undesired components.
The output contains only the desired components.

$$Good/Bad/Ugly \quad \Longrightarrow \quad \boxed{\quad \textbf{Filter} \quad} \quad \longrightarrow \quad Good$$

We are primarily concerned with filters as used in **signal processing.**

The selection principle for these is generally based on the **frequency of the signal components.**

In many cases the input consists of a low-frequency (LF) signal and high-frequency (HF) noise.

The information in the signal can be isolated by using a low-pass filter.

⋆ ⋆ ⋆

In many cases the input consists of a low-frequency (LF) signal and high-frequency (HF) noise.

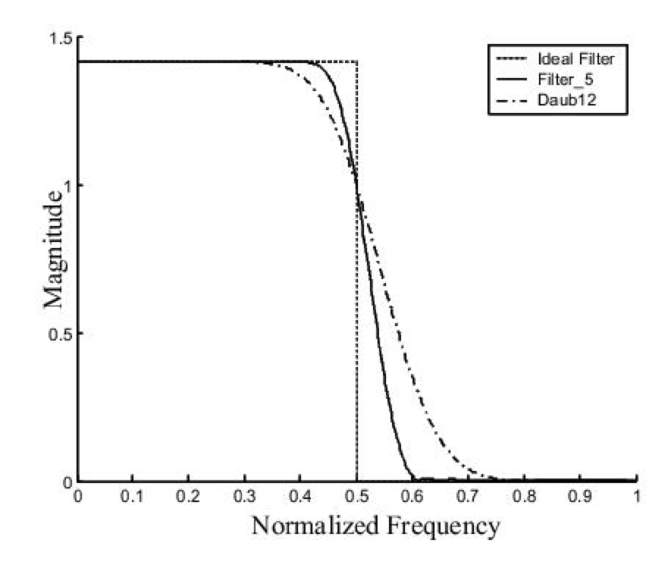
The information in the signal can be isolated by using a low-pass filter.

⋆ ⋆ ⋆

Other ideal filters can be discussed:

- High-pass filters
- Band-pass filters
- Notch filters

The Low-Pass Filter is the one for initialization.



Frequency response of ideal low-pass filter.

Nonrecursive Filters

Given a discrete function of time, $\{x_n\}$, a **nonrecursive digital filter** is defined by

$$y_n = \sum_{k=-N}^{N} a_k x_{n-k}.$$

Nonrecursive Filters

Given a discrete function of time, $\{x_n\}$, a **nonrecursive digital filter** is defined by

$$y_n = \sum_{k=-N}^{N} a_k x_{n-k} .$$

The **output** y_n depends on both past and future values of the **input** x_n.

It does not depend on previous output values.

Recursive Filters

A **recursive** digital filter is defined by

$$y_n = \sum_{k=K}^{N} a_k x_{n-k} + \sum_{k=1}^{L} b_k y_{n-k}$$

where L and N are positive integers. Usually, $K = 0$.

Recursive Filters

A recursive digital filter is defined by

$$y_n = \sum_{k=K}^{N} a_k x_{n-k} + \sum_{k=1}^{L} b_k y_{n-k}$$

where L and N are positive integers. Usually, $K = 0$.

The output y_n at time $n\Delta t$ depends on past and present values of the input and also on previous output values.

Recursive filters are more powerful than non-recursive ones, but the *feedback of the output* can cause numerical instability.

Recursive filters are more powerful than non-recursive ones, but the *feedback of the output* can cause numerical instability.

The response of a nonrecursive filter to an impulse $\delta(n)$ is zero for $|n| > N$, giving rise to the alternative name finite impulse response or FIR filter.

Recursive filters are more powerful than non-recursive ones, but the *feedback of the output* can cause numerical instability.

The response of a nonrecursive filter to an impulse $\delta(n)$ is zero for $|n| > N$, giving rise to the alternative name finite impulse response or FIR filter.

The response of a recursive filter may persist: it is called an infinite impulse response (IIR) filter.

$\star \quad \star \quad \star$

To find the frequency response of a recursive filter, let

$$x_n = \exp(in\theta)$$

and assume an output of the form

$$y_n = H(\theta)\exp(in\theta)$$

To find the **frequency response** of a recursive filter, let

$$x_n = \exp(in\theta)$$

and assume an output of the form

$$y_n = H(\theta)\exp(in\theta)$$

Substitute $y_n = H(\theta)\exp(in\theta)$ into the defining formula

$$y_n = \sum_{k=K}^{N} a_k x_{n-k} + \sum_{k=1}^{L} b_k y_{n-k}$$

To find the frequency response of a recursive filter, let

$$x_n = \exp(in\theta)$$

and assume an output of the form

$$y_n = H(\theta)\exp(in\theta)$$

Substitute $y_n = H(\theta)\exp(in\theta)$ into the defining formula

$$y_n = \sum_{k=K}^{N} a_k x_{n-k} + \sum_{k=1}^{L} b_k y_{n-k}$$

The transfer function $H(\theta)$ is

$$H(\theta) = \frac{\sum_{k=K}^{N} a_k e^{-ik\theta}}{1 - \sum_{k=1}^{L} b_k e^{-ik\theta}}.$$

Again, the transfer function $H(\theta)$ is

$$H(\theta) = \frac{\displaystyle\sum_{k=K}^{N} a_k e^{-ik\theta}}{1 - \displaystyle\sum_{k=1}^{L} b_k e^{-ik\theta}}.$$

Again, the transfer function $H(\theta)$ is

$$H(\theta) = \frac{\sum_{k=K}^{N} a_k e^{-ik\theta}}{1 - \sum_{k=1}^{L} b_k e^{-ik\theta}}.$$

For FIR's, the denominator reduces to unity:

$$H(\theta) = \sum_{k=-N}^{N} a_k e^{-ik\theta}$$

Response function of a FIR:

$$H(\theta) = \sum_{k=-N}^{N} a_k e^{-ik\theta}$$

This equation gives the response once the filter coefficients a_k have been specified.

However, what is really required is the opposite: to choose coefficients that yield the desired response.

This inverse problem has no unique solution, and a great variety of techniques have been developed.

Response function of a FIR:

$$H(\theta) = \sum_{k=-N}^{N} a_k e^{-ik\theta}$$

This equation gives the response once the filter coefficients a_k have been specified.

However, what is really required is **the opposite**: to choose coefficients that yield the desired response.

This **inverse problem** has no unique solution, and a great variety of techniques have been developed.

The entire area of **filter design** is concerned with finding filters having desired properties.

Design of Nonrecursive Filters [skip]

Consider a function of time, $f(t)$, with low and high frequency components.

To filter out the high frequencies:

1. Calculate the Fourier transform $F(\omega)$ of $f(t)$;
2. Set the coefficients of the high frequencies to zero;
3. Calculate the inverse transform.

Design of Nonrecursive Filters [skip]

Consider a function of time, $f(t)$, with low and high frequency components.

To filter out the high frequencies:

1. Calculate the Fourier transform $F(\omega)$ of $f(t)$;
2. Set the coefficients of the high frequencies to zero;
3. Calculate the inverse transform.

Step [2] may be performed by multiplying $F(\omega)$ by an appropriate weighting function $H_c(\omega)$.

Design of Nonrecursive Filters [skip]

Consider a function of time, $f(t)$, with low and high frequency components.

To filter out the high frequencies:

1. Calculate the Fourier transform $F(\omega)$ of $f(t)$;
2. Set the coefficients of the high frequencies to zero;
3. Calculate the inverse transform.

Step [2] may be performed by multiplying $F(\omega)$ by an appropriate weighting function $H_c(\omega)$.

Typically, $H_c(\omega)$ is a step function with cutoff ω_c:

$$H_c(\omega) = \begin{cases} 1, & |\omega| \leq |\omega_c| \\ 0, & |\omega| > |\omega_c| \end{cases}$$

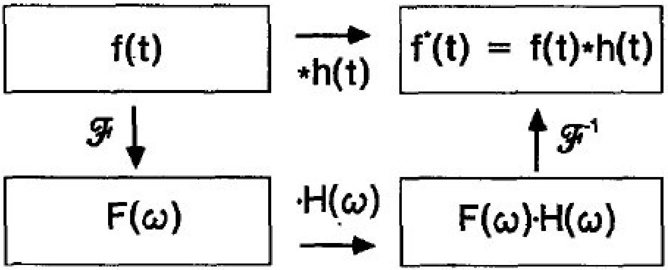


FIG. 1. Schematic representation of the equivalence between convolution and filtering in Fourier space.

Equivalence of filtering and convolution.

$$(h * f)(t) = \mathcal{F}^{-1}\{\mathcal{F}\{h\} \cdot \mathcal{F}\{f\}\}$$

These three steps are equivalent to a **convolution of** $f(t)$ with the inverse Fourier transform of $H_c(\omega)$:

$$h(t) = \sin(\omega_c t)/\pi t\,.$$

This follows from the **convolution theorem**

$$\mathcal{F}\{(h * f)(t)\} = \mathcal{F}\{h\} \cdot \mathcal{F}\{f\} = H_c(\omega) \cdot F(\omega)$$

These three steps are equivalent to a **convolution of** $f(t)$ with the inverse Fourier transform of $H_c(\omega)$:

$$h(t) = \sin(\omega_c t)/\pi t \, .$$

This follows from the **convolution theorem**

$$\mathcal{F}\{(h * f)(t)\} = \mathcal{F}\{h\} \cdot \mathcal{F}\{f\} = H_c(\omega) \cdot F(\omega)$$

Thus, to filter $f(t)$ one calculates

$$f^\star(t) = (h * f)(t) = \int_{-\infty}^{+\infty} h(\tau) f(t-\tau) d\tau.$$

These three steps are equivalent to a **convolution of** $f(t)$ with the inverse Fourier transform of $H_c(\omega)$:

$$h(t) = \sin(\omega_c t)/\pi t\,.$$

This follows from the **convolution theorem**

$$\mathcal{F}\{(h * f)(t)\} = \mathcal{F}\{h\} \cdot \mathcal{F}\{f\} = H_c(\omega) \cdot F(\omega)$$

Thus, to filter $f(t)$ one calculates

$$f^\star(t) = (h * f)(t) = \int_{-\infty}^{+\infty} h(\tau) f(t-\tau) d\tau.$$

For simple functions $f(t)$, this integral may be evaluated analytically. In general, some approximation must be used.

Suppose now that f is known only at **discrete moments** $t_n = n\Delta t$, so that the sequence $\{\cdots, f_{-2}, f_{-1}, f_0, f_1, f_2, \cdots\}$ is given.

Suppose now that f is known only at **discrete moments** $t_n = n\Delta t$, so that the sequence $\{\cdots, f_{-2}, f_{-1}, f_0, f_1, f_2, \cdots\}$ is given.

For example, f_n could be the value of some model variable at a particular grid point at time t_n.

Suppose now that f is known only at discrete moments $t_n = n\Delta t$, so that the sequence $\{\cdots, f_{-2}, f_{-1}, f_0, f_1, f_2, \cdots\}$ is given.

For example, f_n could be the value of some model variable at a particular grid point at time t_n.

The shortest period component which can be represented with a time step Δt is $\tau_{\mathrm{Ny}} = 2\Delta t$, corresponding to a maximum frequency, the so-called Nyquist frequency, $\omega_{\mathrm{Ny}} = \pi/\Delta t$.

The sequence $\{f_n\}$ may be regarded as the Fourier coefficients of a function $F(\theta)$:

$$F(\theta) = \sum_{n=-\infty}^{\infty} f_n e^{-in\theta},$$

where $\theta = \omega\Delta t$ is the digital frequency and $F(\theta)$ is periodic with period 2π: $F(\theta) = F(\theta + 2\pi)$.

[Note: $\theta_{\mathrm{Ny}} = \omega_{\mathrm{Ny}}\Delta t = \pi$]

The sequence $\{f_n\}$ may be regarded as the Fourier coefficients of a function $F(\theta)$:

$$F(\theta) = \sum_{n=-\infty}^{\infty} f_n e^{-in\theta},$$

where $\theta = \omega\Delta t$ is the digital frequency and $F(\theta)$ is periodic with period 2π: $F(\theta) = F(\theta + 2\pi)$.

[Note: $\theta_{\mathrm{Ny}} = \omega_{\mathrm{Ny}}\Delta t = \pi$]

High frequency components of the sequence may be eliminated by multiplying $F(\theta)$ by a function $H_d(\theta)$:

$$H_d(\theta) = \begin{cases} 1, & |\theta| \leq |\theta_c|; \\ 0, & |\theta| > |\theta_c|, \end{cases}$$

The sequence $\{f_n\}$ may be regarded as the Fourier coefficients of a function $F(\theta)$:

$$F(\theta) = \sum_{n=-\infty}^{\infty} f_n e^{-in\theta},$$

where $\theta = \omega \Delta t$ is the digital frequency and $F(\theta)$ is periodic with period 2π: $F(\theta) = F(\theta + 2\pi)$.

[Note: $\theta_{\mathrm{Ny}} = \omega_{\mathrm{Ny}} \Delta t = \pi$]

High frequency components of the sequence may be eliminated by multiplying $F(\theta)$ by a function $H_d(\theta)$:

$$H_d(\theta) = \begin{cases} 1, & |\theta| \leq |\theta_c|; \\ 0, & |\theta| > |\theta_c|, \end{cases}$$

The cutoff frequency $\theta_c = \omega_c \Delta t$ is assumed to fall in the Nyquist range $(-\pi, \pi)$ and $H_d(\theta)$ has period 2π.

The function $H_d(\theta)$ may be expanded:

$$H_d(\theta) = \sum_{n=-\infty}^{\infty} h_n e^{-in\theta} \qquad ; \qquad h_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\theta) e^{in\theta} d\theta.$$

The function $H_d(\theta)$ may be expanded:

$$H_d(\theta) = \sum_{n=-\infty}^{\infty} h_n e^{-in\theta} \qquad ; \qquad h_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\theta) e^{in\theta} d\theta.$$

The values of the coefficients h_n follow immediately:

$$h_n = \frac{\sin n\theta_c}{n\pi}.$$

The function $H_d(\theta)$ may be expanded:

$$H_d(\theta) = \sum_{n=-\infty}^{\infty} h_n e^{-in\theta} \qquad ; \qquad h_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\theta) e^{in\theta} d\theta .$$

The values of the coefficients h_n follow immediately:

$$h_n = \frac{\sin n\theta_c}{n\pi} .$$

Exercise: Prove this.

$\star \quad \star \quad \star$

The function $H_d(\theta)$ may be expanded:

$$H_d(\theta) = \sum_{n=-\infty}^{\infty} h_n e^{-in\theta} \qquad ; \qquad h_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\theta) e^{in\theta} d\theta.$$

The values of the coefficients h_n follow immediately:

$$h_n = \frac{\sin n\theta_c}{n\pi}.$$

Exercise: Prove this.

$\star \quad \star \quad \star$

Let $\{f_n^\star\}$ denote the low frequency part of $\{f_n\}$, with all components having frequency greater than θ_c removed.

The function $H_d(\theta)$ may be expanded:

$$H_d(\theta) = \sum_{n=-\infty}^{\infty} h_n e^{-in\theta} \qquad ; \qquad h_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\theta) e^{in\theta} d\theta.$$

The values of the coefficients h_n follow immediately:

$$h_n = \frac{\sin n\theta_c}{n\pi}.$$

Exercise: Prove this.

$\star \quad \star \quad \star$

Let $\{f_n^\star\}$ denote the low frequency part of $\{f_n\}$, with all components having frequency greater than θ_c removed.

Clearly,

$$H_d(\theta) \cdot F(\theta) = \sum_{n=-\infty}^{\infty} f_n^\star e^{-in\theta}.$$

The **convolution theorem** now implies that $H_d(\theta) \cdot F(\theta)$ is the transform of the convolution of $\{h_n\}$ with $\{f_n\}$:

$$f_n^\star = (h * f)_n = \sum_{k=-\infty}^{\infty} h_k f_{n-k}.$$

The **convolution theorem** now implies that $H_d(\theta) \cdot F(\theta)$ is the transform of the convolution of $\{h_n\}$ with $\{f_n\}$:

$$f_n^\star = (h * f)_n = \sum_{k=-\infty}^{\infty} h_k f_{n-k}.$$

This enables the filtering to be performed in the time domain, i.e., **directly on the sequence** $\{f_n\}$.

$\star \quad \star \quad \star$

The **convolution theorem** now implies that $H_d(\theta) \cdot F(\theta)$ is the transform of the convolution of $\{h_n\}$ with $\{f_n\}$:

$$f_n^\star = (h * f)_n = \sum_{k=-\infty}^{\infty} h_k f_{n-k} .$$

This enables the filtering to be performed in the time domain, i.e., **directly on the sequence** $\{f_n\}$.

$\star \quad \star \quad \star$

In practice the summation must be **truncated**: An approximation to the LF part of $\{f_n\}$ is given by

$$f_n^\star = \sum_{k=-N}^{N} h_k f_{n-k} .$$

The **convolution theorem** now implies that $H_d(\theta) \cdot F(\theta)$ is the transform of the convolution of $\{h_n\}$ with $\{f_n\}$:

$$f_n^\star = (h * f)_n = \sum_{k=-\infty}^{\infty} h_k f_{n-k}.$$

This enables the filtering to be performed in the time domain, i.e., **directly on the sequence** $\{f_n\}$.

$\star \quad \star \quad \star$

In practice the summation must be **truncated**: An approximation to the LF part of $\{f_n\}$ is given by

$$f_n^\star = \sum_{k=-N}^{N} h_k f_{n-k}.$$

We see that the finite approximation to the discrete convolution is identical to a **nonrecursive digital filter**.

Gibbs Oscillations & Window Functions

Truncation of a Fourier series gives rise to **Gibbs oscillations.**

These may be greatly reduced by means of an appropriately defined “window” function.

Gibbs Oscillations & Window Functions

Truncation of a Fourier series gives rise to **Gibbs oscillations.**

These may be greatly reduced by means of an appropriately defined "window" function.

The response of the filter is improved if h_n is multiplied by the **Lanczos window:**

$$w_n = \frac{\sin\bigl(n\pi/(N+1)\bigr)}{n\pi/(N+1)}.$$

Gibbs Oscillations & Window Functions

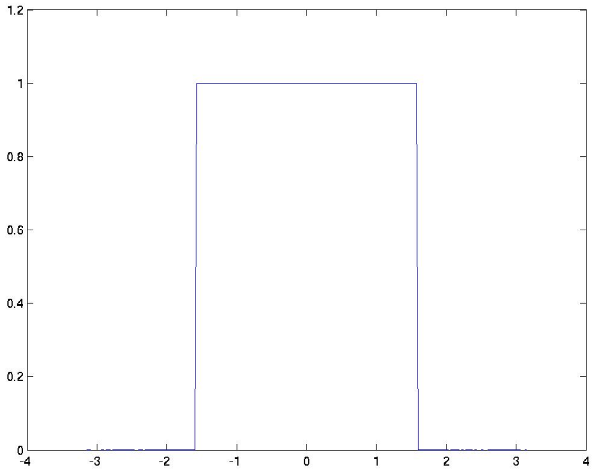
Truncation of a Fourier series gives rise to **Gibbs oscillations.**

These may be greatly reduced by means of an appropriately defined "window" function.

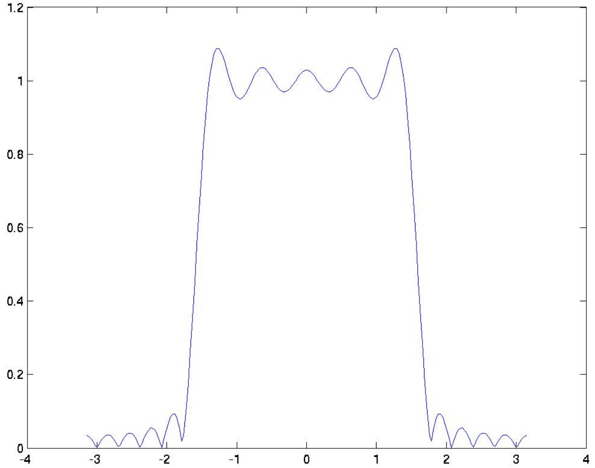
The response of the filter is improved if h_n is multiplied by the **Lanczos window:**

$$w_n = \frac{\sin\big(n\pi/(N+1)\big)}{n\pi/(N+1)}.$$

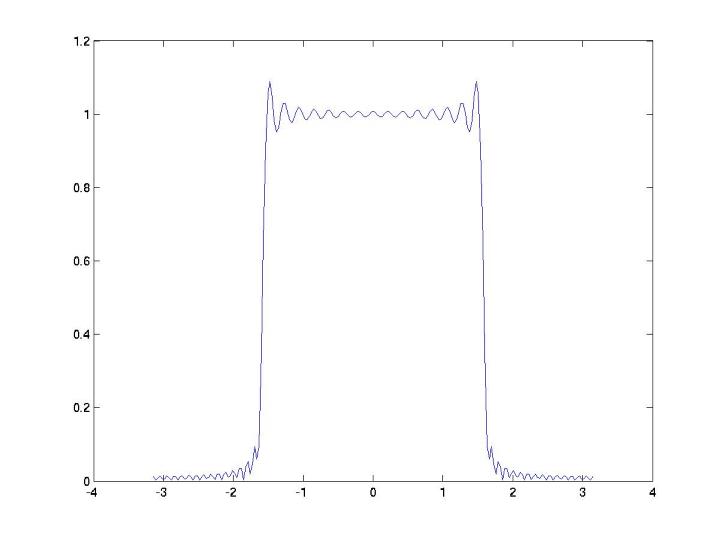
The truncated Fourier analysis of a square wave is shown in the following figures.



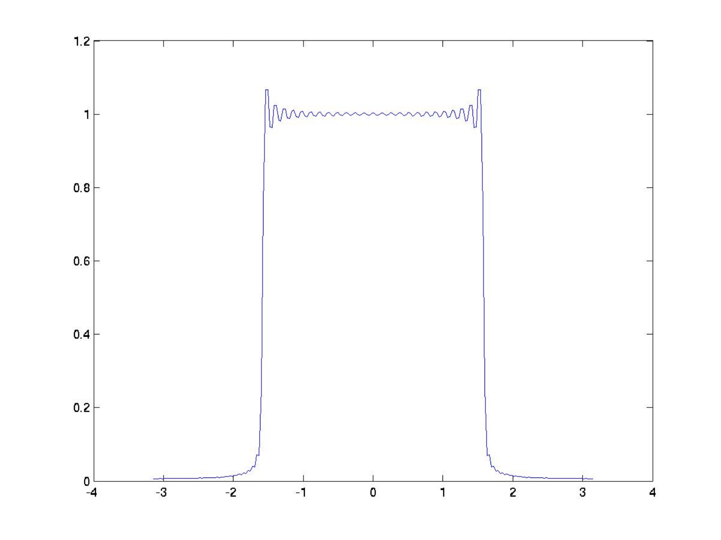
Original Square wave function.



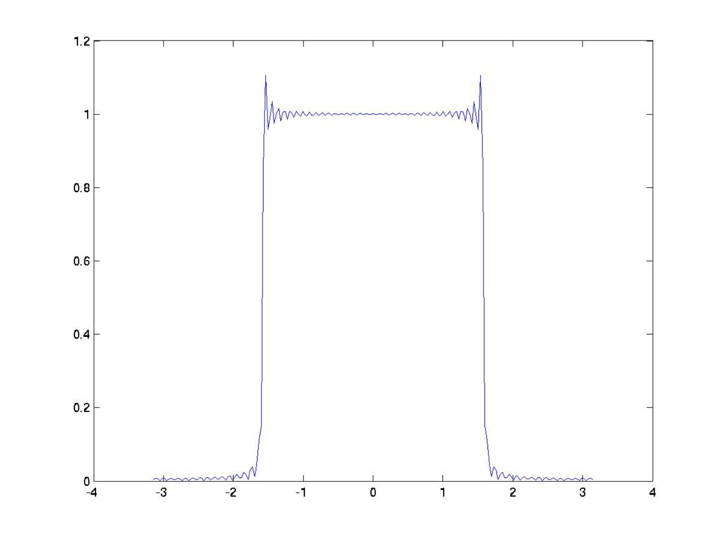
Truncation: $N = 11$ $(N_{\max} = 50)$



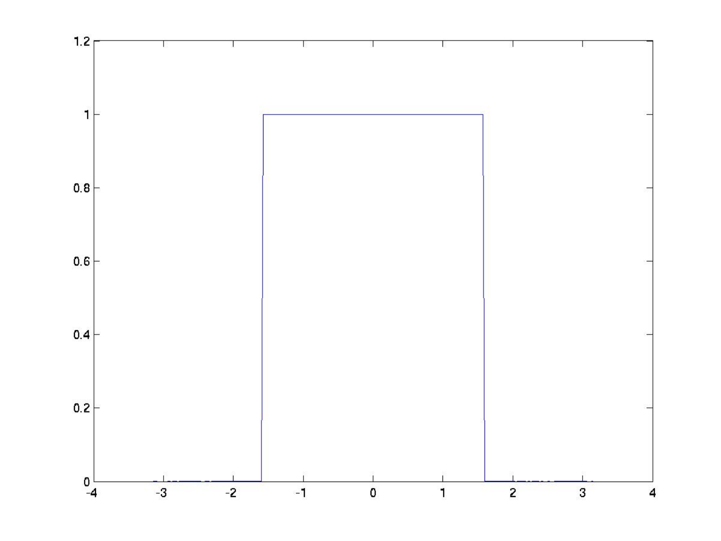
Truncation: $N = 21$ ($N_{\max} = 50$)



Truncation: $N = 31$ ($N_{\text{max}} = 50$)



Truncation: $N = 41$ ($N_{\max} = 50$)



Original Square wave function.

Application of FIR to HiRLAM

An initialization scheme using a nonrecursive digital filter was developed by Lynch and Huang (1992).

Application of FIR to HiRLAM

An initialization scheme using a nonrecursive digital filter was developed by Lynch and Huang (1992).

The value chosen for the cutoff frequency corresponded to a period $\tau_c = 6\ \mathrm{hours}$.

With the time step $\Delta t = 6$ minutes, this corresponds to a (digital) **cutoff frequency** $\theta_c = \pi/30$.

Application of FIR to HiRLAM

An initialization scheme using a nonrecursive digital filter was developed by Lynch and Huang (1992).

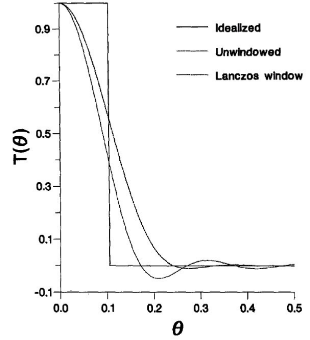
The value chosen for the cutoff frequency corresponded to a period $\tau_c = 6\ \mathrm{hours}$.

With the time step $\Delta t = 6$ minutes, this corresponds to a (digital) cutoff frequency $\theta_c = \pi/30$.

The coefficients were derived by Fourier expansion, truncated at $N = 30$, with a Lanczos window:

$$h_n = \left[\frac{\sin\big(n\pi/(N+1)\big)}{n\pi/(N+1)} \right] \left(\frac{\sin(n\theta_c)}{n\pi} \right) .$$

Application of FIR to HiRLAM



The use of the window decreases the Gibbs oscillations in the stop-band $|\theta| > |\theta_c|$.

However, it also has the effect of widening the pass-band beyond the nominal cutoff.

Application of FIR to HiRLAM

The central lobe of the coefficient function spans a period of six hours, from $t = -3$ h to $t = +3$ h: $T_{\text{Span}} = 6$ **hours.**

The filter summation was calculated over this range, with the coefficients normalized to have unit sum over the span.

Thus, the application of the technique involved computation equivalent to sixty time steps, or a **six hour integration.**

$\star \quad \star \quad \star$

Application of FIR to HiRLAM

The model was first integrated forward for three hours, and running sums of the form

$$f_F^\star(0) = \frac{1}{2} h_0 f_0 + \sum_{n=1}^{N} h_{-n} f_n,$$

where $f_n = f(n\Delta t)$, were calculated for each field (surface pressure, temperature, humidity and winds) at each gridpoint and on each model level.

Application of FIR to HiRLAM

The model was first integrated forward for three hours, and running sums of the form

$$f_F^\star(0) = \frac{1}{2}h_0 f_0 + \sum_{n=1}^{N} h_{-n} f_n,$$

where $f_n = f(n\Delta t)$, were calculated for each field (surface pressure, temperature, humidity and winds) at each gridpoint and on each model level.

These were stored at the end of the 3 hr forecast.

The original fields were then used to make a three hour ‘hindcast’, during which running sums

$$f_B^\star(0) = \frac{1}{2}h_0 f_0 + \sum_{n=-1}^{-N} h_{-n} f_n$$

were computed for each field, and stored as before.

The original fields were then used to make a **three hour 'hindcast'**, during which running sums

$$f_B^\star(0) = \frac{1}{2}h_0 f_0 + \sum_{n=-1}^{-N} h_{-n} f_n$$

were computed for each field, and stored as before.

The two sums were then combined to give

$$f^\star(0) = f_F^\star(0) + f_B^\star(0) = \sum_{n=-N}^{N} h_{-n} f_n \, .$$

The original fields were then used to make a three hour 'hindcast', during which running sums

$$f_B^\star(0) = \frac{1}{2} h_0 f_0 + \sum_{n=-1}^{-N} h_{-n} f_n$$

were computed for each field, and stored as before.

The two sums were then combined to give

$$f^\star(0) = f_F^\star(0) + f_B^\star(0) = \sum_{n=-N}^{N} h_{-n} f_n \, .$$

These fields correspond to the application of the digital filter to the original data.

They are the filtered data.

Phase Errors

In the foregoing, only the amplitudes of the transfer functions have been discussed.

Since these functions are complex, there is also a phase change induced by the filters.

We will not consider this question here. However, it is essential that the phase characteristics of a filter be studied before it is considered for use.

Phase Errors

In the foregoing, only the amplitudes of the transfer functions have been discussed.

Since these functions are complex, there is also a phase change induced by the filters.

We will not consider this question here. However, it is essential that the phase characteristics of a filter be studied before it is considered for use.

Ideally, the phase-error should be as small as possible for the low frequency components which are meteorologically important.

Recall that phase-errors are amongst the most prevalent and pernicious problems in forecasting.

The Dolph-Chebyshev Filter

We now consider a particularly simple filter, having explicit expressions for its impulse response coefficients.

The Dolph-Chebyshev Filter

We now consider a particularly simple filter, having explicit expressions for its impulse response coefficients.

We give here the definition and principal properties of the Dolph-Chebyshev filter.

For further information, see Lynch, 1997
(http://maths.ucd.ie/~plynch)

We use the **Chebyshev polynomials**, defined by

$$T_n(x) = \begin{cases} \cos(n\cos^{-1} x), & |x| \leq 1; \\ \cosh(n\cosh^{-1} x), & |x| > 1. \end{cases}$$

We use the **Chebyshev polynomials**, defined by

$$T_n(x) = \begin{cases} \cos(n \cos^{-1} x), & |x| \leq 1; \\ \cosh(n \cosh^{-1} x), & |x| > 1. \end{cases}$$

Clearly, $T_0(x) = 1$ and $T_1(x) = x$.

We use the **Chebyshev polynomials**, defined by

$$T_n(x) = \begin{cases} \cos(n \cos^{-1} x), & |x| \leq 1; \\ \cosh(n \cosh^{-1} x), & |x| > 1. \end{cases}$$

Clearly, $T_0(x) = 1$ and $T_1(x) = x$.

A recurrence relation follows from the definition:

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), \quad n \geq 2.$$

We use the **Chebyshev polynomials**, defined by

$$T_n(x) = \begin{cases} \cos(n \cos^{-1} x), & |x| \le 1; \\ \cosh(n \cosh^{-1} x), & |x| > 1. \end{cases}$$

Clearly, $T_0(x) = 1$ and $T_1(x) = x$.

A recurrence relation follows from the definition:

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), \qquad n \ge 2.$$

In the interval $|x| \le 1$, $T_n(x)$ varies between $+1$ and -1.

Now consider the function defined in the frequency domain:

$$H(\theta) = \frac{T_{2M}\left(x_0 \cos\left(\theta/2\right)\right)}{T_{2M}(x_0)},$$

where $x_0 > 1$.

Now consider the function defined in the frequency domain:

$$H(\theta) = \frac{T_{2M}\left(x_0 \cos\left(\theta/2\right)\right)}{T_{2M}(x_0)} ,$$

where $x_0 > 1$.

Let θ_s be such that $x_0 \cos(\theta_s/2) = 1$:

- As θ varies from 0 to θ_s, $H(\theta)$ falls from 1 to $r = 1/T_{2M}(x_0)$.
- For $\theta_s \le \theta \le \pi$, $H(\theta)$ oscillates in the range $\pm r$.

Now consider the function defined in the frequency domain:

$$H(\theta) = \frac{T_{2M}\left(x_0 \cos\left(\theta/2\right)\right)}{T_{2M}(x_0)} ,$$

where $x_0 > 1$.

Let θ_s be such that $x_0 \cos(\theta_s/2) = 1$:

- As θ varies from 0 to θ_s, $H(\theta)$ falls from 1 to $r = 1/T_{2M}(x_0)$.
- For $\theta_s \leq \theta \leq \pi$, $H(\theta)$ oscillates in the range $\pm r$.

The form of $H(\theta)$ is that of a low-pass filter with a cut-off at $\theta = \theta_s$.

Now consider the function defined in the frequency domain:

$$H(\theta) = \frac{T_{2M}\left(x_0 \cos\left(\theta/2\right)\right)}{T_{2M}(x_0)}\,,$$

where $x_0 > 1$.

Let θ_s be such that $x_0 \cos(\theta_s/2) = 1$:

- As θ varies from 0 to θ_s, $H(\theta)$ falls from 1 to $r = 1/T_{2M}(x_0)$.
- For $\theta_s \le \theta \le \pi$, $H(\theta)$ oscillates in the range $\pm r$.

The form of $H(\theta)$ is that of a low-pass filter with a cut-off at $\theta = \theta_s$.

$H(\theta)$ can be written as a finite expansion:

$$H(\theta) = \sum_{n=-M}^{+M} h_n \exp(-in\theta).$$

$$H(\theta) = \sum_{n=-M}^{+M} h_n \exp(-in\theta).$$

The coefficients $\{h_n\}$ may be evaluated from the inverse Fourier transform

$$h_n = \frac{1}{N}\left[1 + 2r\sum_{m=1}^{M} T_{2M}\left(x_0 \cos\frac{\theta_m}{2}\right)\cos m\theta_n\right],$$

where $|n| \leq M$, $N = 2M+1$ and $\theta_m = 2\pi m/N$.

$$H(\theta) = \sum_{n=-M}^{+M} h_n \exp(-in\theta).$$

The coefficients $\{h_n\}$ may be evaluated from the inverse Fourier transform

$$h_n = \frac{1}{N}\left[1 + 2r\sum_{m=1}^{M} T_{2M}\left(x_0 \cos\frac{\theta_m}{2}\right)\cos m\theta_n\right],$$

where $|n| \leq M$, $N = 2M+1$ and $\theta_m = 2\pi m/N$.

Since $H(\theta)$ is real and even, h_n are also real and $h_{-n} = h_n$.

$$H(\theta) = \sum_{n=-M}^{+M} h_n \exp(-in\theta).$$

The coefficients $\{h_n\}$ may be evaluated from the inverse Fourier transform

$$h_n = \frac{1}{N}\left[1 + 2r\sum_{m=1}^{M} T_{2M}\left(x_0 \cos\frac{\theta_m}{2}\right)\cos m\theta_n\right] ,$$

where $|n| \leq M$, $N = 2M + 1$ and $\theta_m = 2\pi m/N$.

Since $H(\theta)$ is real and even, h_n are also real and $h_{-n} = h_n$.

The weights $\{h_n : -M \leq n \leq +M\}$ define the Dolph-Chebyshev or, for short, Dolph filter.

In the HiRLAM model, the filter order $N = 2M + 1$ is determined by the time step Δt and forecast span T_S.

In the HiRLAM model, the filter order $N = 2M + 1$ is determined by the time step Δt and forecast span T_S.

The desired frequency cut-off is specified by choosing a value for the cut-off period, τ_s.

Then $\theta_s = 2\pi\Delta t/\tau_s$ and the parameters x_0 and r are

$$\frac{1}{x_0} = \cos\frac{\theta_s}{2}, \qquad \frac{1}{r} = \cosh\left(2M\cosh^{-1}x_0\right) .$$

In the HiRLAM model, the filter order $N = 2M + 1$ is determined by the time step Δt and forecast span T_S.

The desired frequency cut-off is specified by choosing a value for the cut-off period, τ_s.

Then $\theta_s = 2\pi\Delta t/\tau_s$ and the parameters x_0 and r are

$$\frac{1}{x_0} = \cos\frac{\theta_s}{2}, \qquad \frac{1}{r} = \cosh\left(2M\cosh^{-1} x_0\right) .$$

The ripple ratio r is a measure of the maximum amplitude in the stop-band $[\theta_s, \pi]$:

$$r = \left[\frac{\textbf{side-lobe amplitude}}{\textbf{main-lobe amplitude}}\right]$$

In the HiRLAM model, the filter order $N = 2M + 1$ is determined by the time step Δt and forecast span T_S.

The desired frequency cut-off is specified by choosing a value for the cut-off period, τ_s.

Then $\theta_s = 2\pi\Delta t/\tau_s$ and the parameters x_0 and r are

$$\frac{1}{x_0} = \cos\frac{\theta_s}{2}, \qquad \frac{1}{r} = \cosh\left(2M\cosh^{-1}x_0\right) .$$

The ripple ratio r is a measure of the maximum amplitude in the stop-band $[\theta_s, \pi]$:

$$r = \left[\frac{\text{side-lobe amplitude}}{\text{main-lobe amplitude}}\right]$$

The Dolph filter has minimum ripple-ratio for a given main-lobe width and filter order.

Example of Dolph Filter

Suppose components with period less than three hours are to be eliminated ($\tau_s = 3\,\mathrm{h}$) and the time step is $\Delta t = \frac{1}{8}\,\mathrm{h}$.

Example of Dolph Filter

Suppose components with period less than three hours are to be eliminated ($\tau_s = 3\,\mathrm{h}$) and the time step is $\Delta t = \frac{1}{8}\,\mathrm{h}$.

The parameters chosen for the DFI are:

- Span $T_S = 2\,\mathrm{h}$
- Cut-off period $\tau_s = 3\,\mathrm{h}$
- Time step $\Delta t = 450\,\mathrm{s} = \frac{1}{8}\,\mathrm{h}$

Example of Dolph Filter

Suppose components with period less than three hours are to be eliminated ($\tau_s = 3\,\mathrm{h}$) and the time step is $\Delta t = \frac{1}{8}\,\mathrm{h}$.

The parameters chosen for the DFI are:

- Span $T_S = 2\,\mathrm{h}$
- Cut-off period $\tau_s = 3\,\mathrm{h}$
- Time step $\Delta t = 450\,\mathrm{s} = \frac{1}{8}\,\mathrm{h}$

So, $M = 8$, $N = 17$ and $\theta_s = 2\pi\Delta t/\tau_s \approx 0.26$.

Example of Dolph Filter

Suppose components with period less than three hours are to be eliminated ($\tau_s = 3\,\mathrm{h}$) and the time step is $\Delta t = \frac{1}{8}\,\mathrm{h}$.

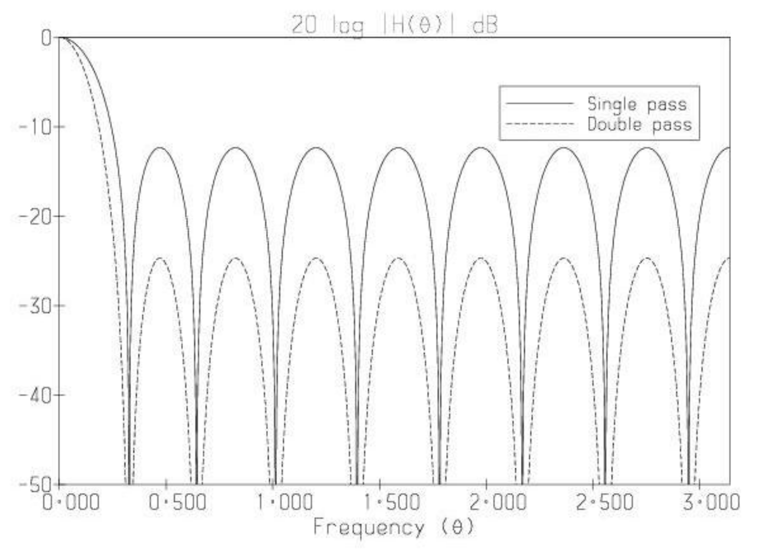
The parameters chosen for the DFI are:

- Span $T_{\mathrm{S}} = 2\,\mathrm{h}$
- Cut-off period $\tau_s = 3\,\mathrm{h}$
- Time step $\Delta t = 450\,\mathrm{s} = \frac{1}{8}\,\mathrm{h}$

So, $M = 8$, $N = 17$ and $\theta_s = 2\pi\Delta t/\tau_s \approx 0.26$.

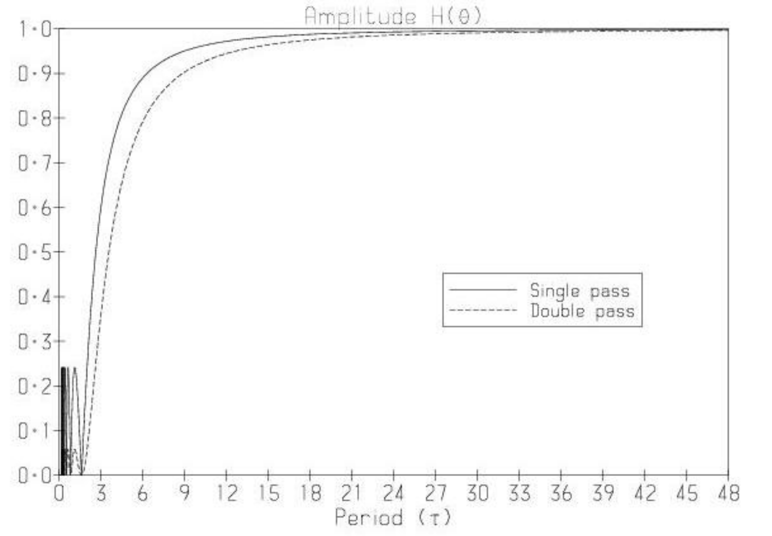
The DFI procedure employed in the HIRLAM model involves a double application of the filter.

We examine the frequency response $H(\theta)$ and its square, $H(\theta)^2$ (a second pass squares the frequency response).



Frequency response for Dolph filter with span $T_S = 2\mathrm{h}$, order $N = 2M + 1 = 17$ and cut-off $\tau_s = 3\mathrm{h}$. Results for single and double application are shown.

Logarithmic response (dB) as a function of frequency.



Frequency response for Dolph filter with span $T_S = 2\mathrm{h}$, order $N = 2M + 1 = 17$ and cut-off $\tau_s = 3\mathrm{h}$. Results for single and double application are shown.

Amplitude response as a function of period.

The ripple ratio of the filter has the value $r = 0.241$.

A single pass attenuates high frequencies (components with $|\theta| > |\theta_s|$) by at least 12.4 dB.

For a double pass, the minimum attenuation is about 25 dB, more than adequate for elimination of HF noise.

The ripple ratio of the filter has the value $r = 0.241$.

A single pass attenuates high frequencies (components with $|\theta| > |\theta_s|$) by at least 12.4 dB.

For a double pass, the minimum attenuation is about 25 dB, more than adequate for elimination of HF noise.

The amplitudes of components with periods less than two hours are reduced to less than 5%.

Components with periods greater than one day are left substantially unchanged.

The ripple ratio of the filter has the value $r = 0.241$.

A single pass attenuates high frequencies (components with $|\theta| > |\theta_s|$) by at least 12.4 dB.

For a double pass, the minimum attenuation is about 25 dB, more than adequate for elimination of HF noise.

The amplitudes of components with periods less than two hours are reduced to less than 5%.

Components with periods greater than one day are left substantially unchanged.

It can be proved (Lynch, 1997) that the Dolph window is an optimal filter whose pass-band edge, θ_p, is the solution of the equation $H(\theta) = 1 - r$.

Implementation in HIRLAM

The digital filter initialization is performed by applying the filter to time series of model variables.

Implementation in HIRLAM

The digital filter initialization is performed by applying the filter to time series of model variables.

The filter is applied in two stages:

In the first stage, a backward integration from $t = 0$ to $t = -T_S$ is performed, with all irreversible physics switched off.

Implementation in HIRLAM

The digital filter initialization is performed by applying the filter to time series of model variables.

The filter is applied in two stages:

In the first stage, a backward integration from $t = 0$ to $t = -T_{\mathrm{S}}$ is performed, with all irreversible physics switched off.

The filter output is calculated by summing:

$$\bar{x} = \sum_{n=0}^{n=-N} h_{N-n} x_n \,.$$

The output $\bar{x}$ is valid at time $t = -\frac{1}{2} T_{\mathrm{S}}$.

In the second stage, a forward integration is made from $t = -\frac{1}{2}T_{\mathrm{S}}$ to $t = +\frac{1}{2}T_{\mathrm{S}}$, starting from the output $\bar{x}$.

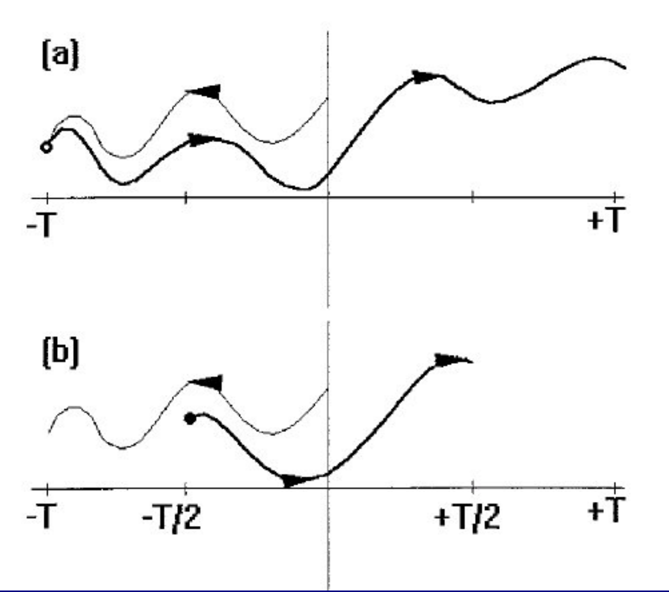
Again, the filter is applied by accumulating sums formally identical to those of the first stage.

In the second stage, a forward integration is made from $t = -\frac{1}{2}T_S$ to $t = +\frac{1}{2}T_S$, starting from the output $\bar{x}$.

Again, the filter is applied by accumulating sums formally identical to those of the first stage.

The output of the second stage is valid at the centre of the interval $[-\frac{1}{2}T_S, +\frac{1}{2}T_S]$, *i.e.*, at $t = 0$.

The output of the second pass is the initialized data.


(a)
-T
+T
(b)
-T
-T/2
+T/2
+T

DFI: Sample Results

The basic measure of noise is the mean absolute value of the surface pressure tendency

$$N_1 = \left(\frac{1}{\mathrm{NGRID}}\right) \sum_{\mathrm{n=1}}^{\mathrm{NGRID}} \left| \frac{\partial p_s}{\partial t} \right| .$$

DFI: Sample Results

The basic measure of noise is the **mean absolute value of the surface pressure tendency**

$$N_1 = \left(\frac{1}{\mathrm{NGRID}}\right)\sum_{n=1}^{\mathrm{NGRID}}\left|\frac{\partial p_s}{\partial t}\right| .$$

For well-balanced fields this quantity has a value of about 1 hPa per 3 hours.

For uninitialized fields, it can be up to an order of magnitude larger.

DFI: Sample Results

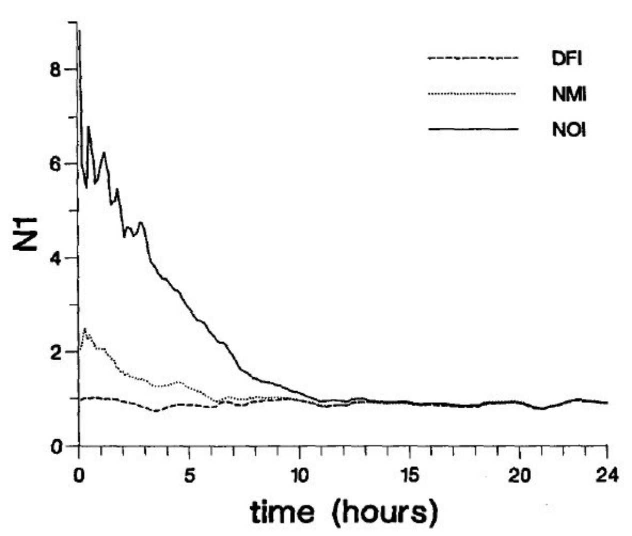
The basic measure of noise is the **mean absolute value of the surface pressure tendency**

$$N_1 = \left(\frac{1}{\mathrm{NGRID}}\right)\sum_{\mathrm{n}=1}^{\mathrm{NGRID}}\left|\frac{\partial p_s}{\partial t}\right| .$$

For well-balanced fields this quantity has a value of about 1 hPa per 3 hours.

For uninitialized fields, it can be up to an order of magnitude larger.

In the following figure, we plot the value of N_1 for three forecasts.



Mean absolute surface pressure tendency for three forecasts. No initialization (NIL); Normal mode initialization (NMI); Digital filter initialization (DFI). (Units hPa/3 h)

The measure N_1 indicates the noise in the vertically integrated divergence field.

However, even when this is small, there may be significant activity in the internal gravity wave modes.

To see this, we look at the vertical velocity field at 500 hPa for the NIL and DFI analyses.

$\star \quad \star \quad \star$

The measure N_1 indicates the noise in the vertically integrated divergence field.

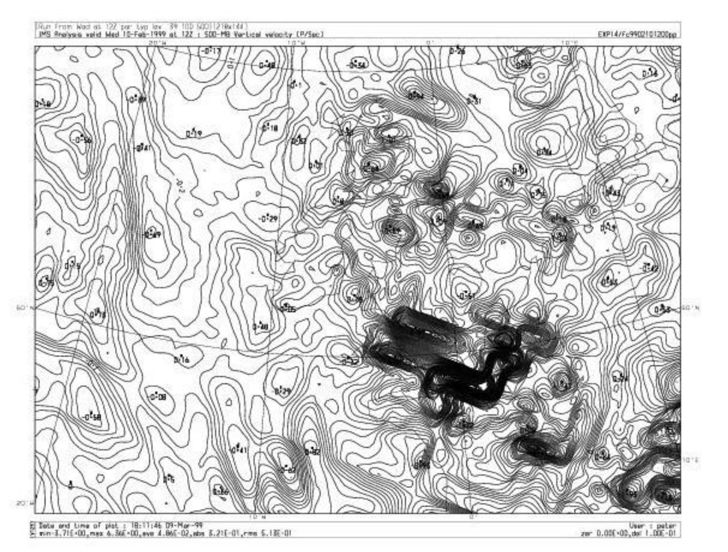
However, even when this is small, there may be significant activity in the internal gravity wave modes.

To see this, we look at the vertical velocity field at 500 hPa for the NIL and DFI analyses.

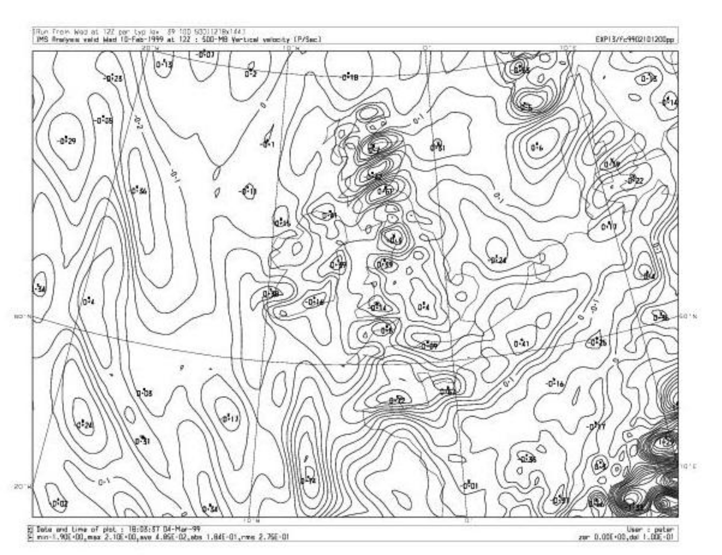
⋆ ⋆ ⋆

The uninitialized vertical velocity field is physically quite unrealistic.

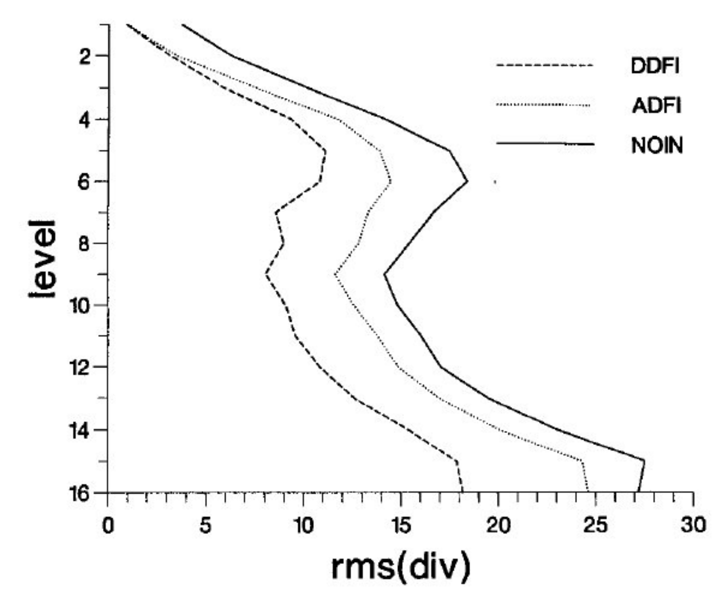
The DFI vertical velocity is much smoother, and much more realistic.



Vertical velocity at 500 hPa for uninitialized analysis (NIL).



Vertical velocity at 500 hPa after digital filtering (DFI).



Root mean square divergence at each model level.

Advantages of DFI

1. No need to compute or store normal modes;
2. No need to separate vertical modes;
3. Complete compatibility with model discretisation;
4. Applicable to exotic grids on arbitrary domains;
5. No iterative numerical procedure to diverge;
6. Ease of implementation and maintenance;
7. Applicable to all prognostic model variables;
8. Applicable to non-hydrostatic models.

Thank you