

§4.3. Digital Filter Initialization

Meteorological analyses must be modified to avoid spurious high frequency oscillations.

The most popular method of initialization up to recently was **nonlinear normal mode initialization**, or NNMI.

This has been widely used, in many NWP centres, and has performed satisfactorily.

However, it has a number of limitations. In particular, it is not straightforward to apply NNMI in **limited geographical domains**.

Recently, an alternative method of initialization, called **digital filter initialization** (DFI), was introduced.

In this lecture we review DFI, and describe how the method is applied in operational NWP.

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 contaminated by **high-frequency (HF) noise**.

The information in the signal can be isolated by using a lowpass filter.

* * *

Other ideal filters can be discussed:

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

But the **Low-Pass Filter** is the one needed for initialization.

The Notion of Filtering

The concept of **filtering** has a rôle in virtually every field of study, “*from topology to theology, seismology to sociology.*”

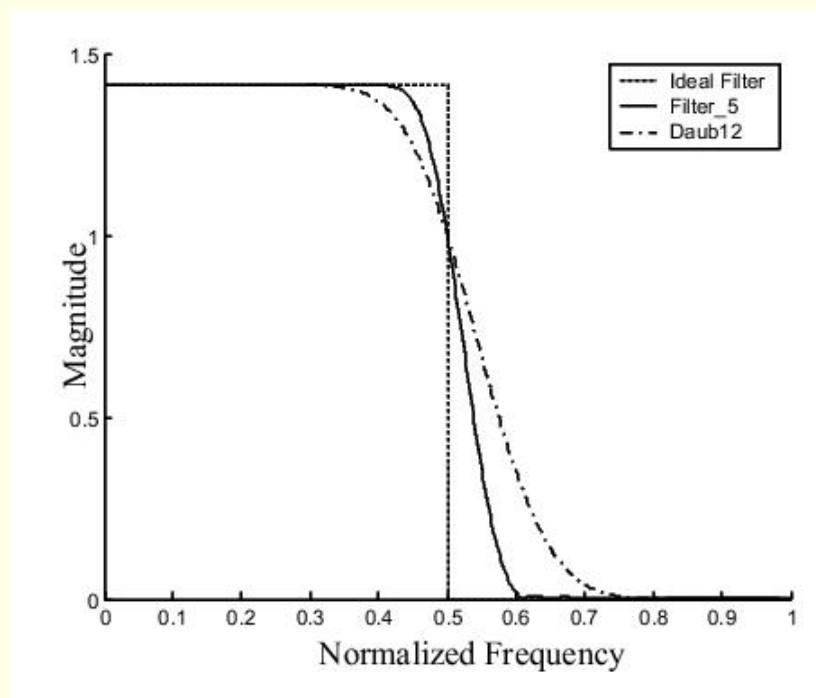
The process of filtering involves the **selection** of those components of an assemblage having some particular property, and the **removal or elimination** of those that lack it.

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

It may be represented by a simple **system diagram**, having an input with both desired and undesired components, and an output comprising only the former.



2



Frequency response of ideal low-pass filter.

Nonrecursive and Recursive 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 at time $n\Delta t$ depends on both past and future values of the **input** x_n , but not on other output values.

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.

5

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.

Since the response of a recursive filter can persist indefinitely, it is called an **infinite impulse response** (IIR) filter.

* * *

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

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

and assume an output of the form

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

6

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

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

For nonrecursive filters 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 derive coefficients which will yield the desired response function.

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

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

To filter out the high frequencies we proceed as follows:

- [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

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

where ω_c is a cutoff frequency.

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

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

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 $\{\dots, f_{-2}, f_{-1}, f_0, f_1, f_2, \dots\}$ is given.

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

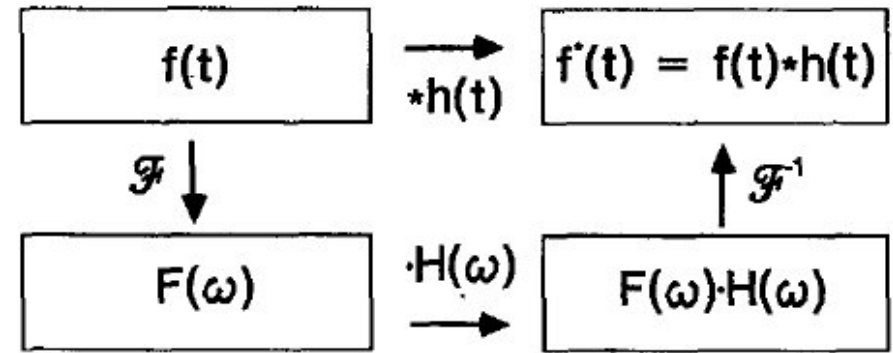


FIG. 1. Schematic representation of the equivalence between convolution and filtering in Fourier space.

Equivalence of **filtering** and **convolution**.

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

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

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

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

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

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:

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

The values of the coefficients h_n follow immediately:

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

* * *

Exercise: Prove this.

* * *

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

Clearly,

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

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

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

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

* * *

In practice the summation must be **truncated**.

Thus, an approximation to the LF part of $\{f_n\}$ is given by

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

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

Gibbs oscillations

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

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

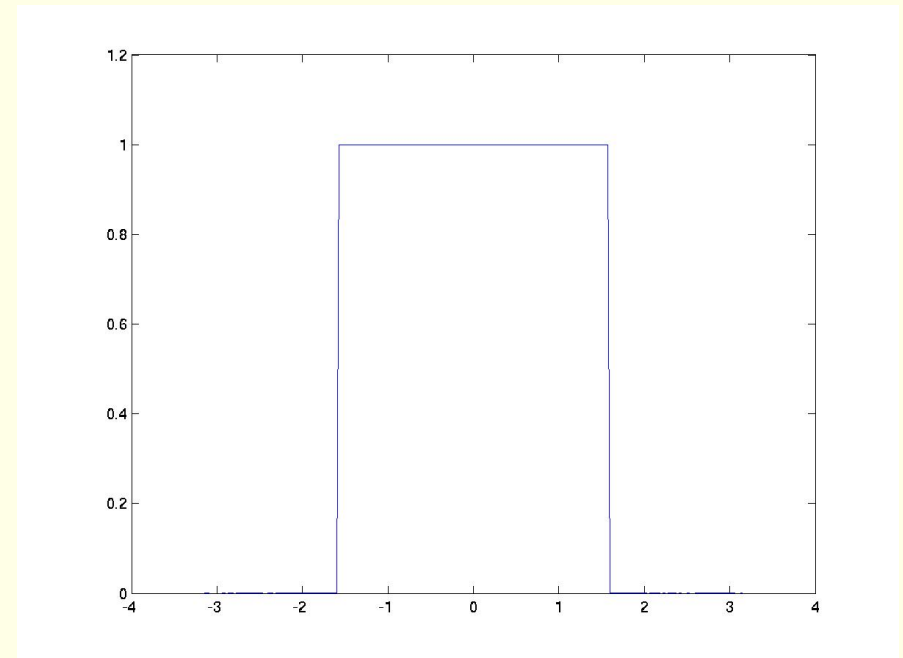
The response of the filter is improved if h_n is multiplied by the **Lanczos window**:

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

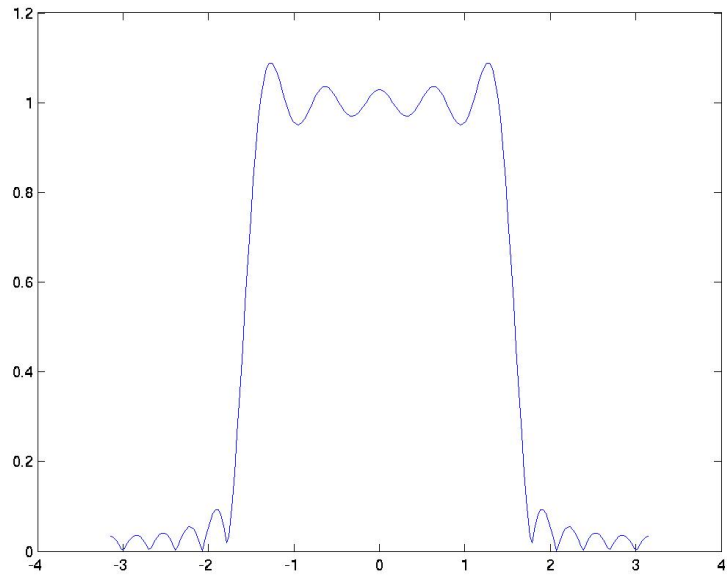
Exercise: Write a MATLAB program to compute the FFT of a step function with various truncations. Thus investigate the Gibbs phenomenon.

* * *

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

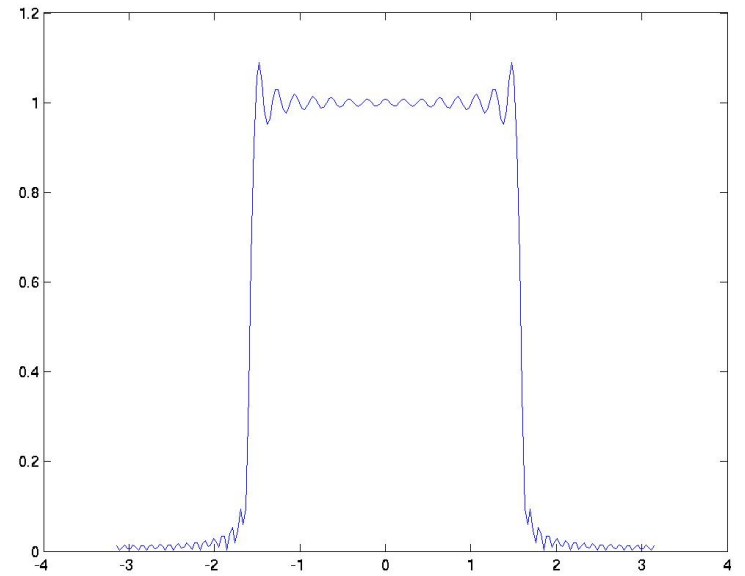


Original Square wave function.



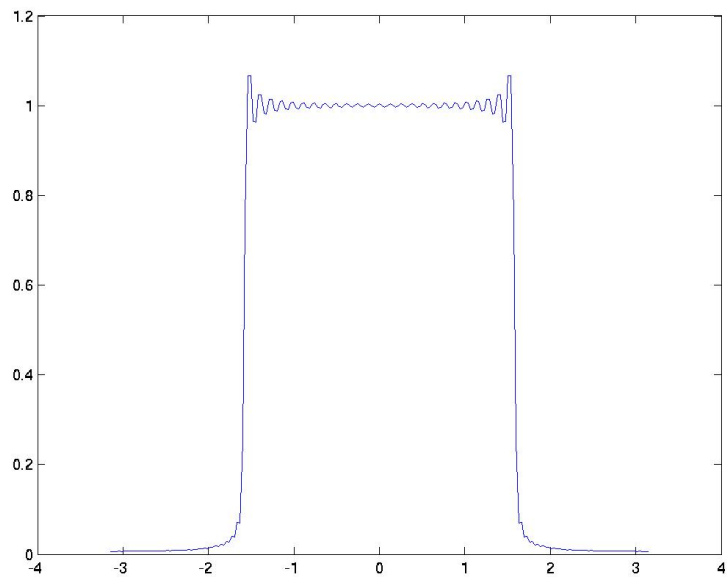
Truncation: $N = 11$ ($N_{\max} = 50$)

17

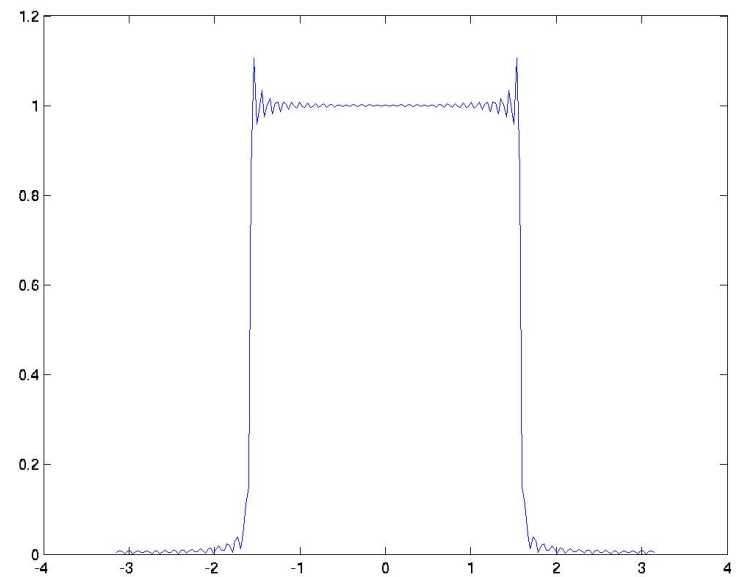


Truncation: $N = 21$ ($N_{\max} = 50$)

18



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



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

Application of FIR to Initialization

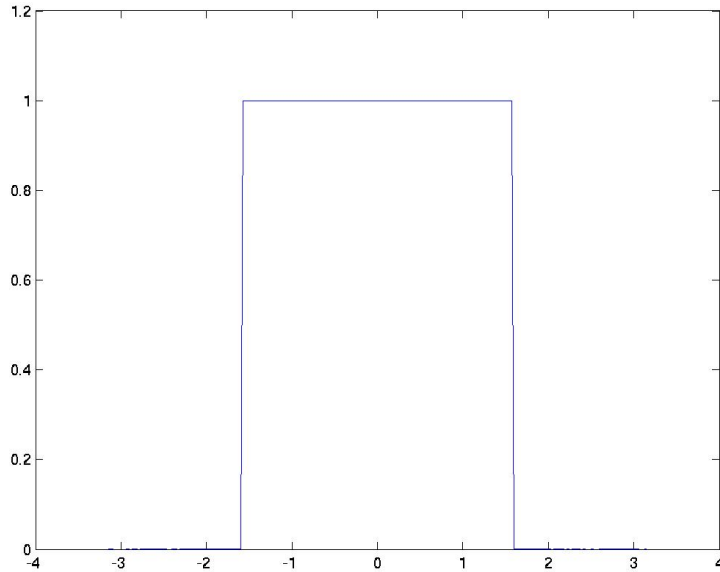
An initialization scheme using a nonrecursive digital filter was developed by Lynch and Huang (1992) for HIRLAM.

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

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

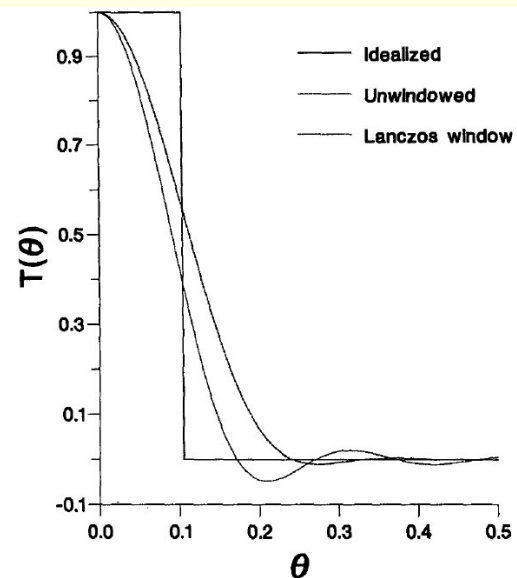
The coefficients were derived by Fourier expansion of a step-function, truncated at $N = 30$, with a Lanczos window:

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



Original Square wave function.

21



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.

For a fuller discussion of windowing see *e.g.* Hamming (1989) or Oppenheim and Schaffer (1989).

22

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**.

* * *

The uninitialized fields of surface pressure, temperature, humidity and winds were **first integrated forward** for three hours, and running sums of the form

$$f_F^*(0) = \frac{1}{2}h_0f_0 + \sum_{n=1}^N h_{-n}f_n,$$

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

These were stored at the end of the three hour forecast.

Repeat:

$$f_F^*(0) = \frac{1}{2}h_0f_0 + \sum_{n=1}^N h_{-n}f_n,$$

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

$$f_B^*(0) = \frac{1}{2}h_0f_0 + \sum_{n=-1}^{-N} h_{-n}f_n$$

were accumulated for each field, and stored as before.

The two sums were then combined to give

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

These fields correspond to the application of the digital filter to the original data. They are **the filtered data**.

25

Break here

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.

It is salutary to recall that **phase-errors are amongst the most prevalent and pernicious problems in forecasting**.

26

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

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

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

A recurrence relation follows from the definition:

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

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

Now consider the function defined in the frequency domain:

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

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 $\theta = \theta_s$.

By means of the definition of $T_n(x)$ and basic trigonometric identities, $H(\theta)$ can be written as a **finite expansion**

$$H(\theta) = \sum_{n=-M}^{+M} h_n \exp(-in\theta).$$

29

Again,

$$H(\theta) = \sum_{n=-M}^{+M} h_n \exp(-in\theta).$$

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**.

30

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 given by

$$\frac{1}{x_0} = \cos \frac{\theta_s}{2}, \quad \frac{1}{r} = \cosh \left(2M \cosh^{-1} x_0 \right).$$

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** ($\tau_s = 3$ h) and the time step is $\Delta t = \frac{1}{8}$ h.

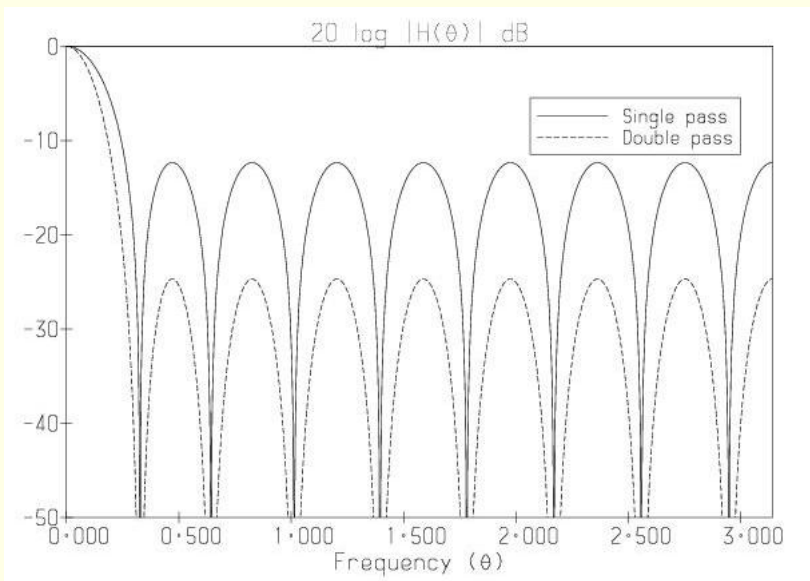
The parameters chosen for the DFI are:

- Span $T_S = 2$ h
- Cut-off period $\tau_s = 3$ h
- Time step $\Delta t = 450$ s = $\frac{1}{8}$ 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**.

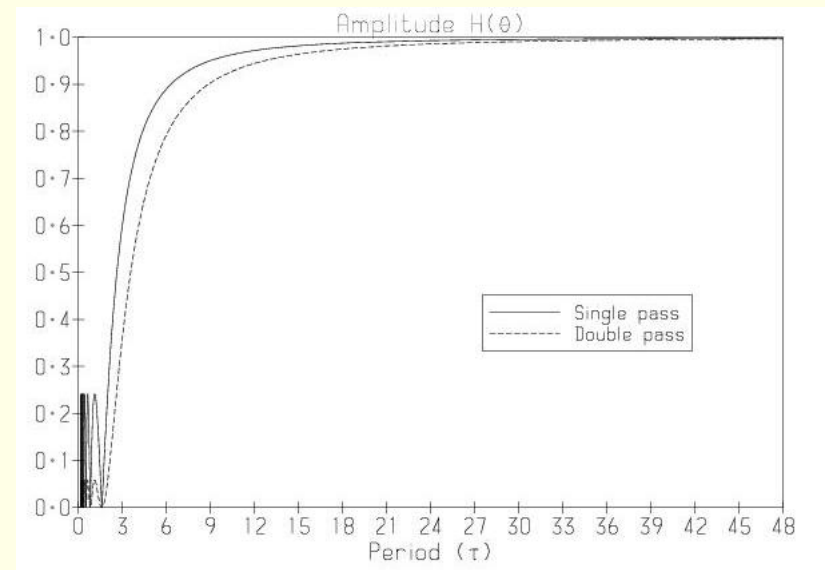
We examine the frequency response $H(\theta)$ and its square, $H(\theta)^2$ (a second pass squares the frequency response).



Frequency response for Dolph filter with span $T_S = 2h$, order $N = 2M + 1 = 17$ and cut-off $\tau_s = 3h$. Results for single and double application are shown.

Logarithmic response (dB) as a function of frequency.

33



Frequency response for Dolph filter with span $T_S = 2h$, order $N = 2M + 1 = 17$ and cut-off $\tau_s = 3h$. Results for single and double application are shown.

Amplitude response as a function of period.

34

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 12.4dB.

For a double pass, the minimum attenuation is about 25dB, more than adequate for elimination of HF noise.

The amplitudes of components with periods less than two hours are reduced to less than 5% of their original value.

Components with periods greater than one day are left substantially unchanged.

It is crucial that an initialization scheme does not distort the meteorologically significant components of the flow.

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**.

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.

The filter output is calculated by accumulating the sums

$$\bar{x} = \sum_{n=0}^{n=-N} h_{N-n} x_n.$$

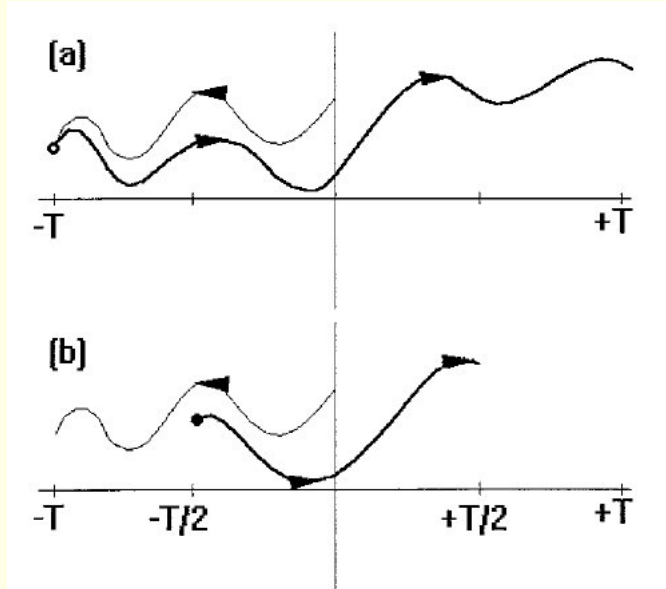
The output \bar{x} is valid at time $t = -\frac{1}{2}T_S$.

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.



37

DFI: Sample Results

The basic measure of noise is the **mean absolute value of the surface pressure tendency**

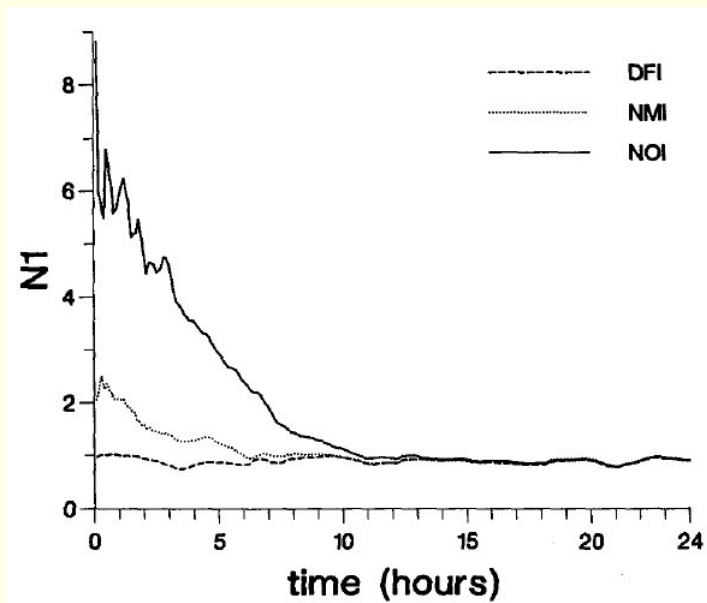
$$N_1 = \left(\frac{1}{\text{NGRID}} \right) \sum_{n=1}^{\text{NGRID}} \left| \frac{\partial p_s}{\partial t} \right|.$$

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.

38



Mean absolute surface pressure tendency for three forecasts. Forecast with no initialization (NIL); normal mode initialization (NMI); digital filter initialization (DFI). Units are hPa/3 hours.

The measure N_1 indicates the noise in the vertically integrated divergence field.

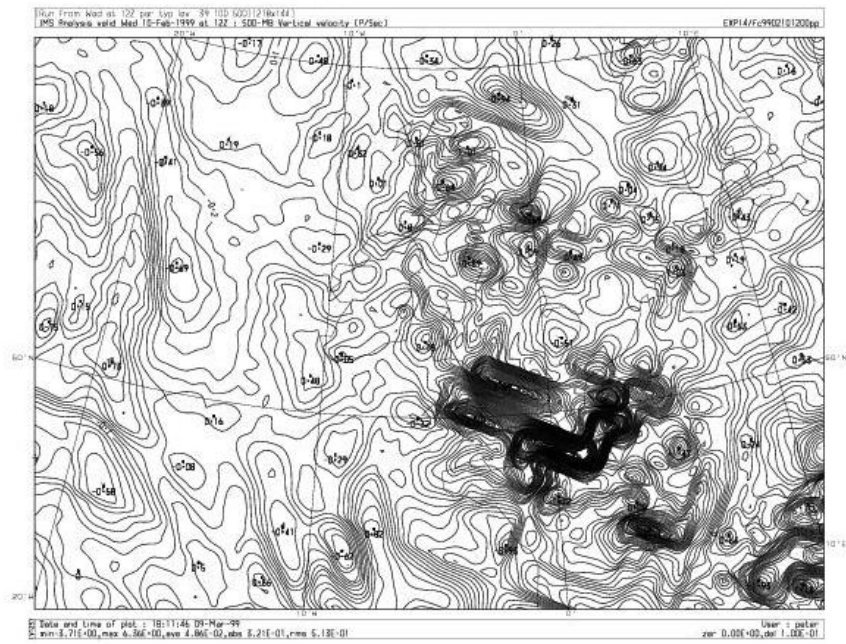
However, even when this is small, there may be significant activity in the **internal gravity wave modes**.

To see this, we look at the vertical velocity field at 500 hPa for the NIL and DFI analyses.

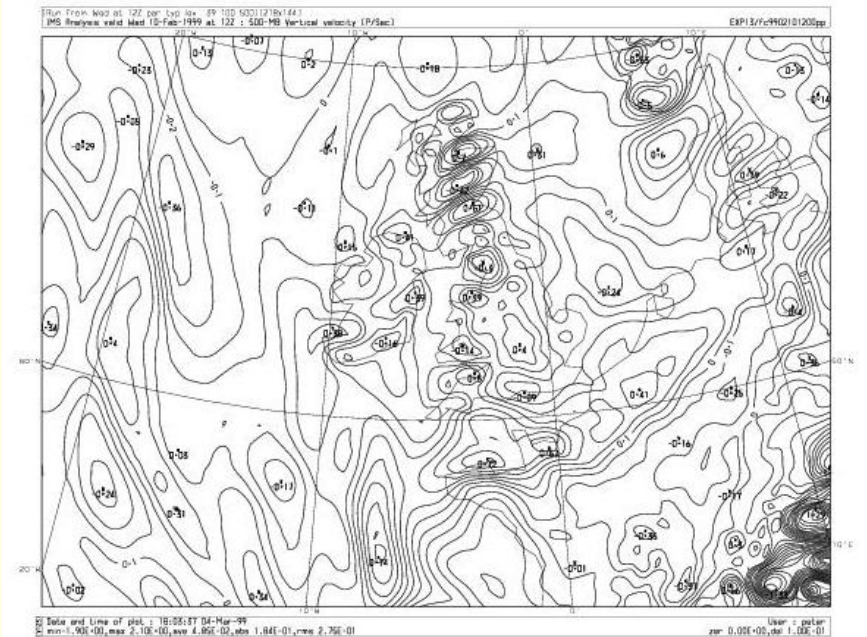
* * *

The uninitialized vertical velocity field is physically quite unrealistic.

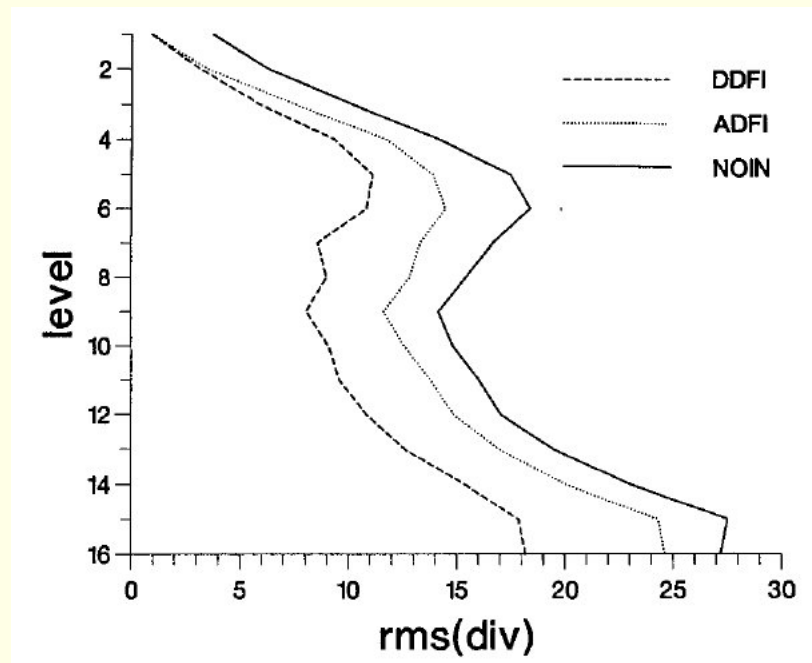
The DFI vertical velocity is much smoother, and much more realistic.



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 which may diverge;
6. Ease of implementation and maintenance;
7. Applicable to all prognostic model variables;
8. Applicable to non-hydrostatic models.

End of §4.3