The University of Reading

Scale Analysis of Observational Information in Data Assimilation

G.M. Baxter, S.L. Dance, A.S. Lawless and N.K. Nichols

NUMERICAL ANALYSIS REPORT 11/2006

Department of Mathematics
The University of Reading
Whiteknights, PO Box 220
Reading
Berkshire RG6 6AX

Department of Mathematics

Abstract

In order to improve our ability to forecast high impact weather we need forecasting models with a high resolution. However, if we want a forecast model with a very high resolution it can only have a limited domain size, due to restrictions in computer power. For example, the Met Office High Resolution Trial Model has a resolution of 1km and a domain of 300x300 grid points and 76 vertical levels. One problem caused by this limited domain size is that the atmosphere can have waves whose wavelength is longer than the domain of the limited area model. To investigate how observations of these waves can be assimilated into the model we consider here what scales the observational information is projected onto. We introduce the discrete Fourier transform (DFT) and investigate how it may be used to understand wave properties. We show how applying the DFT techniques to a simple variational analysis can allow us to identify and investigate the properties of the analysis and in particular enable us to differentiate between phase and amplitude errors. We consider cases with both perfect observations and observations with error.

1 Introduction

Due to advances in computer power, many weather forecasting centres are now developing forecasting models which are able to represent much smaller scales than were previously possible, down to the order of 1km resolution. However, these higher resolutions have some problems. The problems under consideration here stem from the fact that if we want a model with a very high resolution it can only have a limited domain size, due to restrictions in computer power. For example, the Met Office High Resolution Trial Model has a resolution of 1km and a domain of 300x300 grid points and 76 vertical levels. One problem caused by this is that the atmosphere can have waves whose wavelength is longer than the domain of our limited area model and therefore cannot be represented. We can however have observations of all the different wavelengths present, so the question is how to identify the wave being observed and how these 'long waves' can be assimilated into the model. Our aim is to look at what scales the observational information projects onto. In particular we consider

- What happens if observations are taken at every grid point?
- What happens if observations are not taken at every grid point?
- What are the properties of our solutions and how can we identify those properties?
- How can we look at and compare the phase and amplitude of different solutions?
- How can we differentiate between the different errors in our solutions?

In order to include observational data in the model we use data assimilation. To do this we have:

- a background state/prior estimate $\boldsymbol{x^b}$;
- a vector of observations $\boldsymbol{y}$;

- a matrix $\boldsymbol{H}$ which is the linearisation of h, where h is an observation operator from model to observation space. For example in the case of a direct observation it might be the interpolation of a wind observation that is not located at a grid point, for an indirect observation it might be the physical relationships connecting an observed radar reflectivity and the model variables;
- a covariance matrix of background errors $\boldsymbol{B}$;
- a covariance matrix of observation errors $\boldsymbol{R}$;
- an analysis $\boldsymbol{x^a}$, which is what we generate;
- a 'true state' $\boldsymbol{x^t}$, which is what we are trying to estimate.

We use variational data assimilation which means we try to find $\boldsymbol{x^a}$ which minimises a cost function J as a least squares problem [7]. The cost function J is a measure of the distance of the solution from the background and the observations, weighted by the inverse of the error covariance matrices.

In Section 2 we introduce the discrete Fourier transform (DFT) and investigate how it may be used to understand wave properties. In Sections 3 and 4 we apply the DFT techniques to a simple variational analysis and investigate how the DFT can allow us to identify and understand the properties of the analysis and in particular enable us to differentiate between phase and amplitude errors. Perfect observations are used in Section 3 while in Section 4 we use observations with error. Conclusions are drawn in Section 5.

2 The Discrete Fourier Transform

In order to investigate the scales being projected onto we need to introduce the discrete Fourier transform (DFT). The continuous Fourier transform of a function $f(\xi)$ is defined as

$$F(v) = \int_{-\infty}^{\infty} f(\xi) \exp(-2\pi i v \xi) d\xi. \tag{1}$$

where v and ξ are the continuous transform variables. However when the variables are discrete, which they are if digital computers are used to perform the analysis, equation (1) needs to be generalised for a discrete function. This can be done by letting $f_j \equiv f(\xi_j)$, with $j = 0, \cdots, N-1$. This gives the discrete Fourier transform of the function f_j as

$$DFT\,(f_j) = \widetilde{f}_k = \sum_{j=0}^{N-1} f_j e^{-i\xi_j k}, \tag{2}$$

and the inverse transform is

$$f_j = \frac{1}{N} \sum_{k=0}^{N-1} \widetilde{f}_k e^{i\xi_j k}, \tag{3}$$

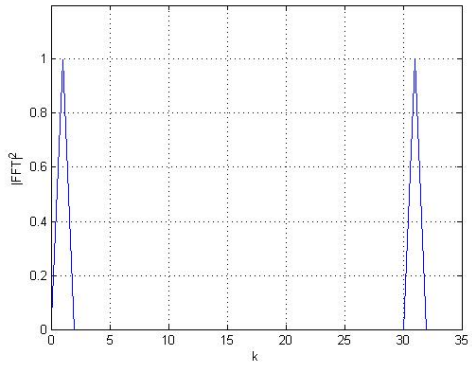
where $\xi_j = \frac{2\pi j}{N}$, $j = 0, \cdots, N-1$, and k is the wavenumber [2].

This discrete Fourier transform (DFT) can be calculated on a computer using a method called the Fast Fourier Transform (FFT). This FFT reduces the number of computer operations needed from $O(N^2)$ to $O(N \lg(N))$ [1]. For small values of N this is not so significant but for large values of N this reduction in computational expense is essential in making the calculations possible.

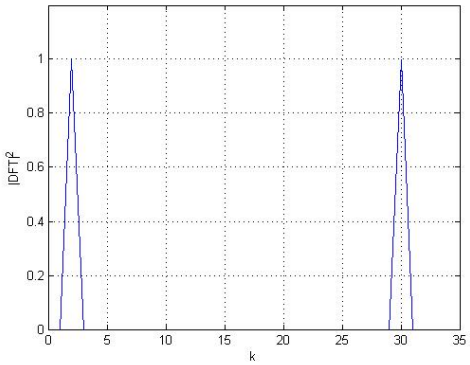
2.1 The Power Spectrum

A convenient way of analysing the DFT coefficients is to consider their complex modulus. A suitably scaled [1] plot of these complex moduli is called a power spectrum [8]. Figure 1(a) shows the power spectrum of $\sin(\xi_j)$. In order to better understand this plot and why it looks the way it does, we consider its different features separately.

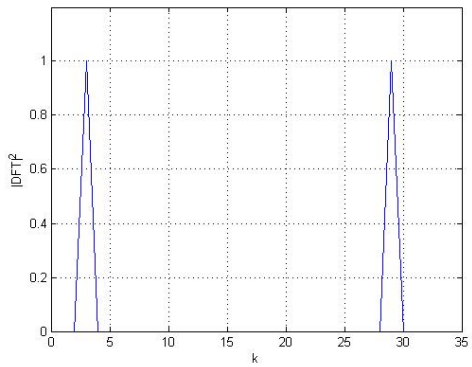
[1] As with all Fourier transforms there is a normalisation factor (M or $1/M$) that can be put on either the DFT or its inverse. Because of this it can be necessary to rescale the outputted DFT coefficients so that the amplitudes are 1 instead of M. In our case the domain is $[0, 2\pi]$, i.e. a double domain, so $M = N/2$ and the rescaling factor needed is $2/N$.



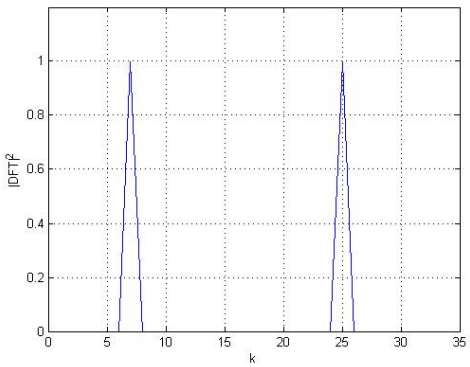
(a) The power spectrum of $\sin(\xi_j)$.



(b) The power spectrum of $\sin(2\xi_j)$.



(c) The power spectrum of $\sin(3\xi_j)$.



(d) The power spectrum of $\sin(7\xi_j)$.

Figure 1: The power spectra for waves with different wavenumbers.

2.1.1 Shape

Firstly, the two peaks that can be seen in Figure 1(a) correspond to a particular property of the discrete Fourier transform; for a real sequence f_j with $j = 0, \cdots, N-1$, its DFT will, in general, be a sequence of N complex numbers. In particular, for real f_j, $\widetilde{f}_k$ and $\widetilde{f}_{N-k}$ are related by

$$\widetilde{f}_{N-k} = \widetilde{f}_k^* \tag{4}$$

for $k = 0, \cdots, N-1$, where $*$ denotes the complex conjugate [2]. This property can also be written as

$$\left|\widetilde{f}_{N-k}\right|^2 = \left|\widetilde{f}_k\right|^2$$

and it is this that causes a periodic function to have transform peaks in two places, not one, as can be seen in figure 1(a). The two peaks correspond to the fact that $\sin(x) = -\sin(2\pi - x)$.

The reason the power spectrum has the shape of these isolated peaks can be better understood by doing the DFT calculations by hand for this simple sine wave example.

2.1.2 Calculating the DFT by Hand

In order to understand better the DFT we do the calculations by hand for the simple case of $f_j = \sin(\xi_j)$. Using this f_j, equation (2) becomes

$$\widetilde{f}_k = \sum_{j=0}^{N-1} \sin(\xi_j) e^{-i\xi_j k}. \tag{5}$$

Writing $\sin(\xi_j) = \frac{1}{2i}\left(e^{i\xi_j} - e^{-i\xi_j}\right)$ this becomes

$$\widetilde{f}_k = \frac{1}{2i} \sum_{j=0}^{N-1} \left(e^{i\xi_j} - e^{-i\xi_j}\right) e^{-i\xi_j k}, \tag{6}$$

which can be rearranged to give

$$\widetilde{f}_k = \frac{1}{2i} \sum_{j=0}^{N-1} e^{-i\xi_j (k-1)} - \frac{1}{2i} \sum_{j=0}^{N-1} e^{-i\xi_j (k+1)}. \tag{7}$$

In order to evaluate equation (7) we use the fact that the complex exponential function obeys the following orthogonality relation.

Lemma 1

$$\frac{1}{N}\sum_{j=0}^{N-1} e^{-is\xi_j} = \begin{cases} 1 & \text{if } s = Nm,\ m \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases}$$

For proof of Lemma 1 see [4].

Using Lemma (1) with $s = k-1$ and then $s = k+1$ and noting that $\frac{1}{i} = -i$, equation (7) becomes

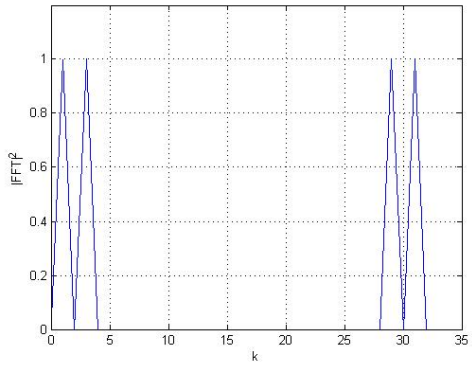
$$\widetilde{f}_k = \begin{cases} \frac{-iN}{2} & k = 1, \\ \frac{iN}{2} & k = N-1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

This agrees with the power spectrum of $\sin(\xi_j)$ plotted in Figure 1(a), which has a peak at $k = 1$ and $k = N-1$ and is zero everywhere else. By applying the normalisation factor $\frac{1}{N/2}$ we also get the amplitude to be one.

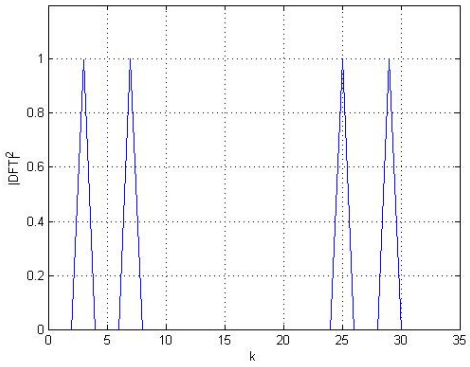
With this very simple example we can begin to see why the DFT looks the way it does. The peak form of the power spectrum is caused by a single non-zero value of the DFT. These calculations also begin to suggest why the peak is located where it is. This will now be considered more closely.

2.1.3 Location

We now consider the location of the peaks. In order to do this we start by again considering the simple case of a basic sine wave, but now we vary the wavenumber κ. The power spectra of $\sin(\kappa x)$, with $\kappa = 1, 2, 3, 7$ are plotted in Figure 1. As can be seen, the location of the peaks varies as the wavenumber changes. By comparing the graphs it can be seen that the location of the peaks depends on the wavenumber. For $\kappa = 1$ the peaks are located at $k = 1, 31$ $(N-1 = 31)$; similarly for $\kappa = 2$ the peaks are at $k = 2, 30$ $(N-2 = 30)$, for $\kappa = 3$ the peaks are at $k = 3, 29$ and for $\kappa = 7$ the peaks are at $k = 7, 25$. In other words the peaks are located at $k = \kappa$ and $k = N-\kappa$, where this second location corresponds to the periodic condition (4). This property of the peak location can also be seen by considering again the calculations done by



(a) The power spectrum of $\sin(\xi_j) + \sin(3\xi_j)$.



(b) The power spectrum of $\sin(3\xi_j) + \sin(7\xi_j)$.

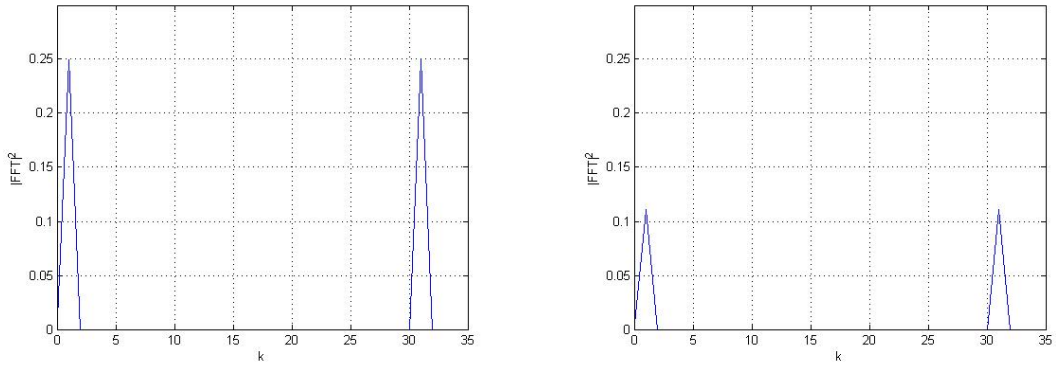
Figure 2: Power spectra for multiple waves.

hand, but changing $\sin(\xi_j)$ to $\sin(\kappa\xi_j)$. With $f_j = \sin(\kappa\xi_j)$, equation (8) becomes

$$\widetilde{f}_k = \begin{cases} \frac{-iN}{2} & k = \kappa, \\ \frac{iN}{2} & k = N - \kappa, \\ 0 & otherwise. \end{cases}$$

This correlation between the wavenumber and the peak location is one of the things that makes the power spectrum so useful. It allows us to use the power spectrum to pick out the waves present in a solution.

In order to see what the power spectrum would look like for multiple waves, or inversely to consider what is indicated by a power spectrum with multiple peaks, we now calculate $y = \sin(x) + \sin(3x)$ and $y = \sin(3x) + \sin(7x)$ and plot their power spectra. As can be seen in Figure 2, in both the $y = \sin(x) + \sin(3x)$ and $y = \sin(3x) + \sin(7x)$ cases, there are two clear sets of peaks. By comparison with Figure 1 we can see that there is one set corresponding to each wavelength. This is because the FFT is a linear transform, so if two or more waveforms are summed to give a new wave then the power spectrum of this new wave will be the sum of the power spectra of the original waves [8]. This shows how the power spectrum can be used as an indicator of the wave pattern, the two sets of peaks indicating that there are two distinct waveforms present. The location of the peaks also indicates the wavenumber of each wave present. This will be the case for wavenumbers $\kappa < N/2$,



(a) The power spectrum of $\frac{1}{2}\sin(\xi_j)$.

(b) The power spectrum of $\frac{1}{3}\sin(\xi_j)$.

Figure 3: Power spectra for waves of different amplitude.

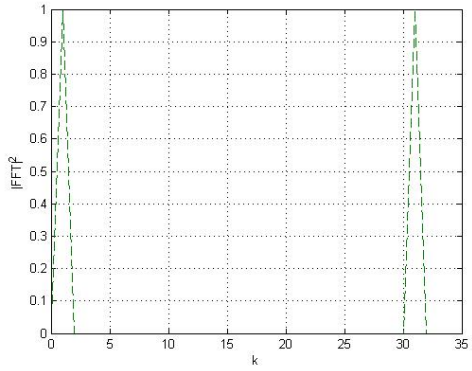
since this is the largest wavenumber (conversely smallest wavelength) that can be represented on a grid with N grid points.

Having considered the location of the peaks we next consider the amplitude.

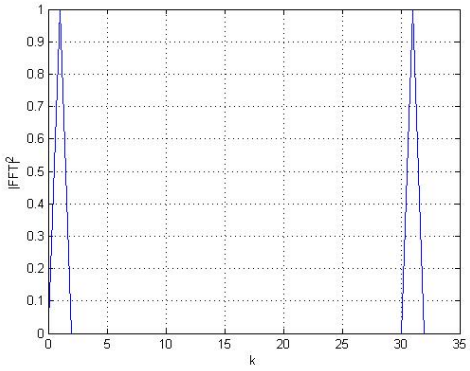
2.1.4 Amplitude

The amplitude of the power spectrum remains unchanged for varying wavenumber, as can be seen in Figure 2. Instead we consider the power spectra of scaled sine waves $y = \alpha\sin(\xi_j)$, where α is a real positive constant. Figure 3 shows the power spectra for $\alpha = \frac{1}{2}$ and $\alpha = \frac{1}{3}$. As can be seen the amplitude of the power spectrum varies as α changes. As the sine wave is scaled by factor α, the amplitude of the power spectrum is scaled by factor α^2, corresponding to the squaring of the FFT when plotting the power spectrum. This property can also be seen by considering the definition of the DFT, equation (2). The scaling factor α applied to the sine function can be brought outside of the summation since it does not depend on j. Therefore α acts as a scaling factor on the whole DFT. The power spectrum can be used to identify and retrieve a scaling factor α from a solution.

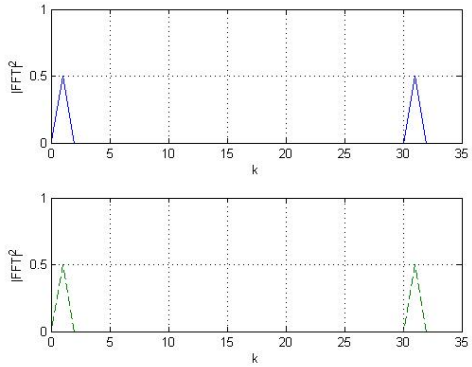
So far we have considered the complex modulus of the DFT. Now we take the real and imaginary parts separately and consider the phase of the wave.



(a) $\sin(\xi_j)$.



(b) $\sin(\xi_j + \frac{\pi}{2})$.



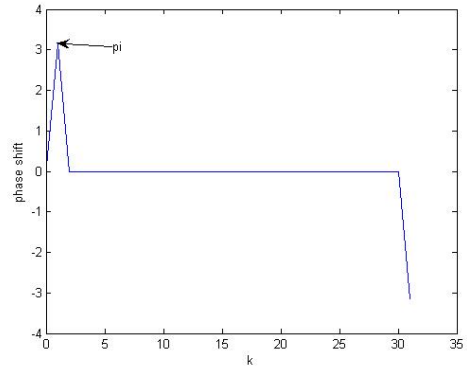
(c) $\sin(\xi_j + \frac{\pi}{4})$.

Figure 4: The solid line shows the real part and the dashed line shows the imaginary part of the power spectrum.

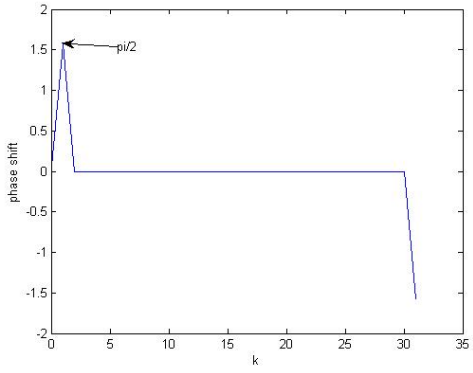
2.1.5 Phase

Figure 4(a) is a plot of the square of the absolute value of the real (solid) and imaginary (dashed) parts of the DFT of the function $f_j = \sin(\xi_j)$. As can be seen, the DFT has purely imaginary coefficients. This is because sine is an odd function. It can be shown from the definition of the DFT that odd functions have purely imaginary DFT coefficients and even functions have purely real DFT coefficients [2]. This second property can be seen by considering an equivalent graph for the function $f_j = \cos(\xi_j)$. The cosine function is equivalent to a sine function with a phase shift of $\pi/2$ and this difference in phase is picked out by the DFT.

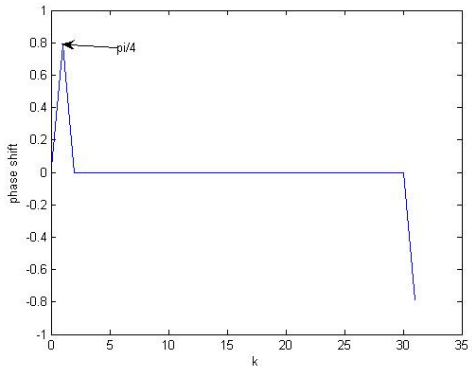
The phase can be considered further by also considering a general phase shift. We take the wave $\sin(\xi_j + \phi)$, where ϕ represents a phase shift. We take $\phi = \pi, \frac{\pi}{2}, \frac{\pi}{4}$ and consider the real and imaginary parts of the DFT for each value of ϕ. When $\phi = \pi$, the real and imaginary power spectra remain unchanged from those of the sine wave



(a) $\phi = \pi$.



(b) $\phi = \frac{\pi}{2}$.



(c) $\phi = \frac{\pi}{4}$.

Figure 5: Phase shift of $\sin(\xi_j + \phi)$ for $\phi =$ (a) π, (b) $\frac{\pi}{2}$ and (c) $\frac{\pi}{4}$.

shown in Figure 4(a). This corresponds to the fact that $\sin(\xi_j + \pi) = -\sin(\xi_j)$, so it is still an odd function. Figure 4(b) shows that when $\phi = \frac{\pi}{2}$, the power spectrum is all in the real part. This is because $\sin(\xi_j + \frac{\pi}{2}) = \cos(\xi_j)$ and cosine is an even function. It can be seen in Figure 4(c) that for $\phi = \frac{\pi}{4}$, half the amplitude is in the real part and half is in the imaginary part. This is logical since the wave is now half way between an odd and an even function.

We can also use the DFT to regain the phase shift ϕ. The phase of a complex number is its argument so by calculating the argument of the DFT, which is a complex number, we have the phase of the wave. We consider now the DFT of $\sin(\xi_j + \phi)$ and also take the DFT of $\sin(\xi_j)$. We calculate the phase of both waves. Where the DFT is zero, the phase is also set to zero. We define the phase shift as $phaseshift = phase(\sin(\xi_j + \phi)) - phase(\sin(\xi_j))$, and plot the results. Figure 5 shows these phase shift plots for our three values of the phase shift ϕ. As can be seen, the height of the peak in each of these plots is equal to the phase shift ϕ. Therefore we can use the DFT and a phase shift plot to regain an unknown phase shift ϕ from a solution.

We have shown that a lot of information about a wave can be learned by simply considering its DFT. It is this that makes the DFT so useful. For a simple sine wave we have shown that wavenumber and pattern, amplitude and phase can all be found from the DFT and its power spectrum. We have also shown methods for how this can be done. We now go on to apply these methods to more complicated waveforms and aim to use them to understand the analysis we obtain from our cost function.

3 The Assimilation Problem

In order to understand better the observational information and onto what scales it is projected, we start by taking a wave with wavelength equal to our model domain size.

In order to investigate the problem, we begin by setting up a simple framework within which to work. We take the 1-D interval $[0, 2\pi)$ and divide it into a grid of N equally spaced points ξ_j, where

$$\xi_j = \frac{2\pi j}{N}, \qquad j \in [0, \ldots, N-1].$$

For all purposes throughout this work, N is taken as $N = 32$, unless otherwise stated.

Now that we have set up a framework within which to work and have tools with which to examine our analysis we can start to consider the assimilation problem.

3.1 The Cost Function

In order to generate our analysis $\boldsymbol{x^a}$ we use a cost function J. Here we start by taking the standard 3D-Var cost function [6]

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x^b})^T \boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{x^b}) + \frac{1}{2}(\boldsymbol{y} - h(\boldsymbol{x}))^T \boldsymbol{R}^{-1}(\boldsymbol{y} - h(\boldsymbol{x})). \tag{9}$$

The cost function J is minimised for $J(\boldsymbol{x} = \boldsymbol{x^a})$. The vector $\boldsymbol{x}$ is often called the 'state vector'.

Although h can be non-linear it can be approximated by linearisation around the background state. We let

$$\boldsymbol{x} = \boldsymbol{x^b} + \boldsymbol{\delta x}, \tag{10}$$

and then

$$h(\boldsymbol{x^b} + \boldsymbol{\delta x}) \approx h(\boldsymbol{x^b}) + \boldsymbol{H}\boldsymbol{\delta x}. \tag{11}$$

The matrix $\boldsymbol{H}$ is the linearisation of h about $\boldsymbol{x^b}$ and is often referred to as the Jacobian.

We now substitute equations (10) and (11) into equation (9) and rearrange slightly. We then minimise J by setting $\nabla_x J = 0$ at the analysis $\boldsymbol{x} = \boldsymbol{x^a}$. We

then rearrange the expression for $\boldsymbol{\delta x^a}$ and using the vector identity

$$(\boldsymbol{B}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H})\boldsymbol{B}\boldsymbol{H}^T = \boldsymbol{H}^T\boldsymbol{R}^{-1}(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T),$$

the solution can be written in the form

$$\boldsymbol{x^a} = \boldsymbol{x^b} + \boldsymbol{B}\boldsymbol{H}^T(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T)^{-1}(\boldsymbol{y} - h(\boldsymbol{x^b})). \tag{12}$$

In operational systems this solution is not used and the 3D-Var cost function is instead minimised iteratively using a descent algorithm [6]. This has the benefit that large matrices (for weather forecasting the state vector can be of size $n \sim 10^7$ meaning matrices with $\sim 10^{14}$ elements) do not have to be inverted. However for our small test problem the dimensions are such that we can use the analytic solution (12).

3.2 A Diagonal Matrix $\boldsymbol{B}$

3.2.1 A Zero Background $\boldsymbol{x^b} = \boldsymbol{0}$

We begin by letting

$$y_j = x_j^t,$$

where y_j is the observation and x_j^t is the truth at grid point j. In other words we assume perfect observations, taken at the grid points. The observation operator h is therefore linear and $h = \boldsymbol{H} = \boldsymbol{\gamma I}$, where

$$\{\gamma_j\} = \begin{cases} 1 & \text{if there is an observation at grid point } j \\ 0 & \text{otherwise.} \end{cases}$$

We assume the error covariance matrices $\boldsymbol{B}$ and $\boldsymbol{R}$ to be diagonal, with background and observation variances σ_b^2 and σ_o^2 respectively. We also assume the background $\boldsymbol{x^b} = 0$. The cost function (9) can therefore be written as

$$J = \sum_j x_j^2\sigma_b^{-2} + \sum_j (y_j - x_j)^2\,\sigma_o^{-2}, \tag{13}$$

which has at its minimum

$$0 = \frac{\partial J}{\partial x} = \sum_j 2x_j\sigma_b^{-2} + \sum_j 2\,(y_j - x_j)\,(-1)\,\sigma_o^{-2},$$

$$0 = 2x_j\sigma_b^{-2} - 2\left(y_j - x_j\right)\sigma_o^{-2},$$

$$0 = 2x_j\sigma_o^2 - 2y_j\sigma_b^2 + 2x_j\sigma_b^2,$$

$$0 = x_j\left(\sigma_o^2 + \sigma_b^2\right) - y_j\sigma_b^2.$$

The analysis is then given by

$$x_j^a = \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2} y_j, \tag{14}$$

for $j = 0, \ldots, N-1$. We assume no time dependence.

To start with we consider an example in which the true solution is

$$x_j^t = \frac{1}{10}\sin(\xi_j).$$

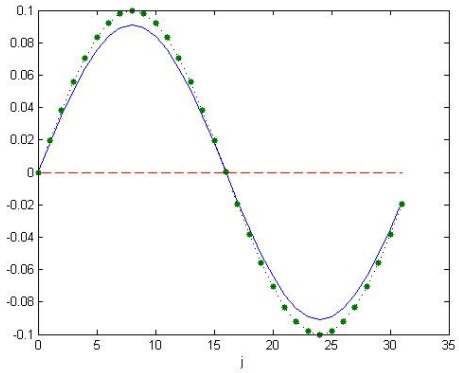
We make this choice because we have assumed a background $\boldsymbol{x^b} = 0$, which is not a particularly good approximation to an ordinary sine wave. We therefore use a scaled sine wave so that its amplitude is smaller and $\boldsymbol{x^b} = 0$ is a better approximation. We take the observation and background variances to be

$$\sigma_o^2 = \frac{1}{10}, \qquad \sigma_b^2 = 1, \tag{15}$$

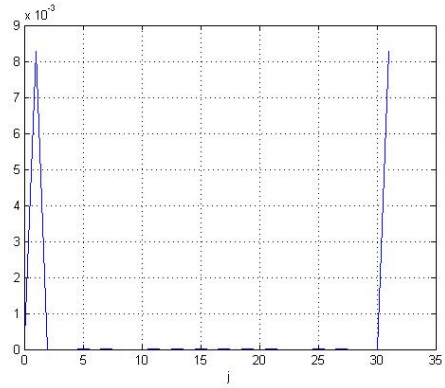
which means the weighting is in favour of the observations. We give the observations more influence than the background over the analysis since we have more faith in the accuracy of our observations.

We now calculate our analysis using equation (14), taking observations at (*i*) every grid point, (*ii*) every other grid point and (*iii*) only two grid points. At grid points where no observation is taken the analysis is set to the background. All observations are perfect observations taken from the truth. In order to understand our analysis and best see what is going on, the analysis vector $\boldsymbol{x^a}$ and a power spectrum for each example are plotted in Figure 6. The power spectrum of $\frac{1}{10}\sin(\xi_j)$ is plotted in Figure 7 for comparison.

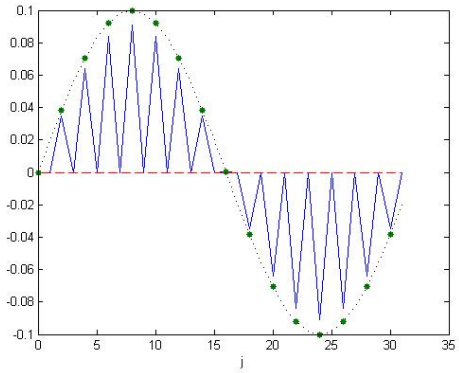
These graphs do show what we expect. Figure 6(a) shows the analysis for example (*i*) and the amplitude of $\boldsymbol{x^a}$ is lower than that of the observations. This is due to the influence of the background $\boldsymbol{x^b} = 0$. As we saw in Section 2.1.4 this scaling factor on the analysis should be visible in the power spectrum as a reduction in amplitude. This agrees with the power spectrum of the analysis shown in Figure 6(b)



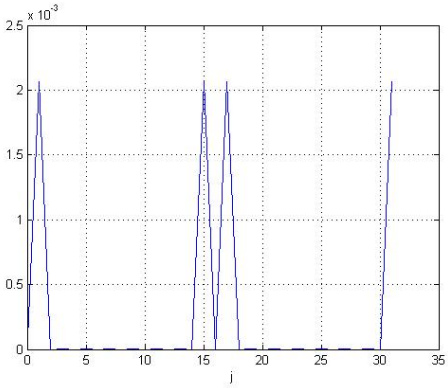
(a) The analysis $\boldsymbol{x}^{\boldsymbol{a}}$ with observations taken every grid point.



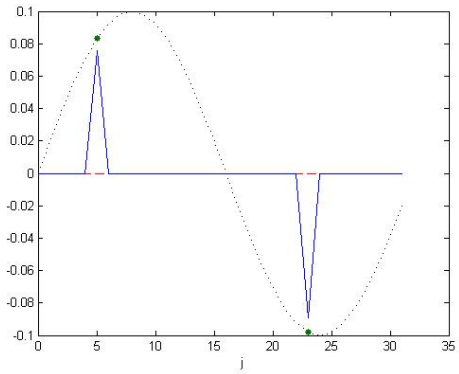
(b) The power spectrum where observations are taken every grid point.



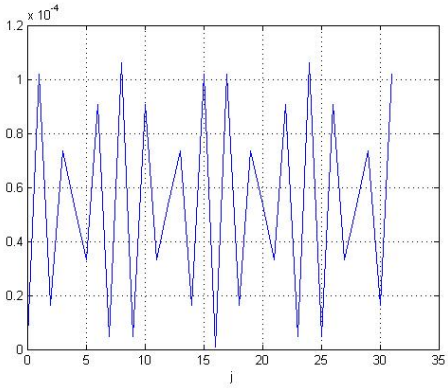
(c) The analysis $\boldsymbol{x}^{\boldsymbol{a}}$ with observations taken every other grid point.



(d) The power spectrum where observations are taken every other grid point.



(e) The analysis $\boldsymbol{x}^{\boldsymbol{a}}$ with observations taken at only two grid points.



(f) The power spectrum where observations are taken at only two grid points.

Figure 6: The analysis $\boldsymbol{x}^{\boldsymbol{a}}$ are plotted as well as the power spectrum for each analysis. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and • show the observations.

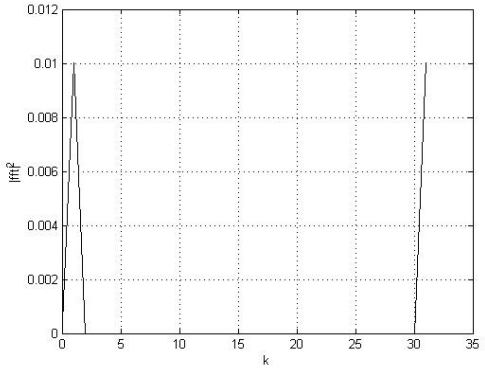


Figure 7: The power spectrum of $\frac{1}{10}\sin(\xi_j)$.

which looks like the power spectrum that we would expect for a sine wave but with a lower peak, as can be seen by comparison with Figure 7.

In example (ii) the analysis $\boldsymbol{x^a}$ is a rapidly oscillating wave, as can be seen in Figure 6(c). This rapid oscillation is caused by having a diagonal error covarience matrix $\boldsymbol{B}$. It is the matrix $\boldsymbol{B}$ which spreads out the observational information. If $\boldsymbol{B}$ is diagonal then an observation at a grid point only affects that grid point; there is no influence on neighbouring points. This lack of spreading in the observational information results here in a rapidly oscillating wave since at every point where no observation is taken the solution is set to the value of the background, which is zero in this example. The wave has two modes, firstly the envelope of the wave corresponds to a wave with wavelength $\lambda = 2\pi$, wavenumber $k = 1$. Secondly the oscillations themselves have a wavelength equal to twice the distance between two grid points, i.e. $\lambda = 2\Delta x$. On a grid $[0, L)$ which is divided into $N = 32$ sections, one grid space Δx has length $\Delta x = \frac{L}{32}$. The wavelengths that can be resolved on this grid are $\frac{L}{1}$ to $\frac{L}{16}$. The longest wavelength that can be resolved $\lambda = \frac{L}{1}$ corresponds to wavenumber $k = 0$, a constant function, and the shortest wavelength that can be resolved $\lambda = \frac{L}{16}$ corresponds to wavenumber $k = 15$. The rapid oscillations in our analysis have wavelength $\lambda = 2\Delta x = \frac{L}{16}$ so have wavenumber $k = 15$. To see what a power spectrum for a wave of wavelength $k = 15$ looks like on our $[0, 2\pi)$ domain, the power spectrum of $\sin(15\xi_j)$ is plotted in Figure 8. By comparing Figures 6(d), 7, and 8 we can see that the two wavelengths present in the analysis are clearly visible in the power spectrum, where there are two sets of peaks, each one corresponding

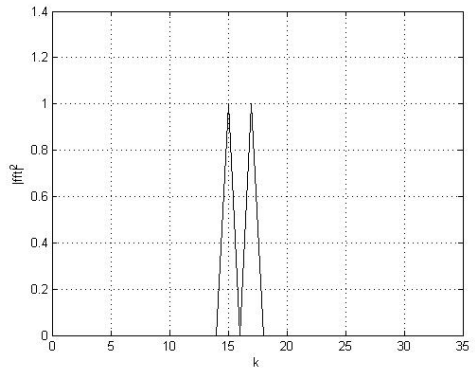


Figure 8: The power spectrum of $\sin(15\xi_j)$.

to the power spectrum of a single mode.

It is also worth noting the reduction in amplitude of the peaks at $k = 1$ in the power spectrum for example (ii), shown in Figure 6(d), where the amplitude is less than a quarter of that shown in Figure 7. This can be understood by again considering Section 2.1.4 and looking more closely at Figure 6(c). As can be seen the analysis oscillates, but it is important also to note that it is oscillating about a wave with a significantly lower amplitude than that of $x_j^t = \frac{1}{10}sin(\xi_j)$, the wave from which the observations are taken. This lower amplitude can be thought of as a scaling factor on the wave envelope and is what causes the power spectrum to have a lower amplitude at $k = 1$. The power spectrum is picking up the scaling factor, just as was shown in Section 2.1.4.

The power spectrum for example (iii) shown in Figure 6(f) does not really reveal a lot, but this is to be expected from looking at Figure 6(e). The analysis $\boldsymbol{x}^a$ can not easily be represented by a wave or a summation of waves. There are no clearly defined wavelengths present and it is not a mixture of identifiable modes, so a meaningful power spectrum would not be expected.

3.2.2 A Non-Zero Background $\boldsymbol{x^b}$

We now take the true solution to be a sine wave with no scaling factor

$$x_j^t = \sin(\xi_j).$$

We still take perfect observations of this true solution. A zero background is not a very good approximation to a sine wave so we now use a non-zero background $\boldsymbol{x^b} = x_j^b$. This affects the cost function as the background term can no longer be neglected and equation (13) becomes

$$J = \sum_j \left(x_j^b - x_j\right)^2 \sigma_b^{-2} + \sum_j \left(y_j - x_j\right)^2 \sigma_o^{-2}, \tag{16}$$

which is minimised by

$$0 = \frac{\partial J}{\partial x} = \sum_j 2\left(x_j^b - x_j\right)(-1)\,\delta_{ij}\sigma_b^{-2} + \sum_j 2\left(y_j - x_j\right)(-1)\,\delta_{ij}\sigma_o^{-2},$$

$$0 = -2\left(x_j^b - x_j\right)\sigma_b^{-2} - 2\left(y_j - x_j\right)\sigma_o^{-2},$$

$$0 = x_j^b\sigma_o^2 + y_j\sigma_b^2 - x_j\left(\sigma_b^2 + \sigma_o^2\right).$$

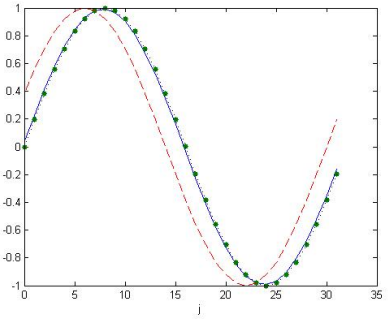
The analysis is therefore given by

$$x_j^a = \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2} y_j + \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2} x_j^b, \tag{17}$$

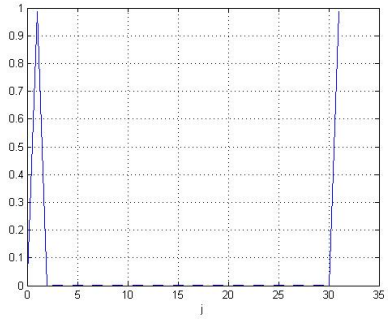
where $j = 0, \ldots, N-1$. As can be seen, the solution now has an extra term compared to that in equation (14). This extra term depends on x_j^b. By setting $\boldsymbol{x^b} = \mathbf{0}$ we get back to the solution we had before.

We keep $\sigma_b^2 = 1$ and $\sigma_o^2 = 0.1$ as before, so emphasis is again placed on the observations. We set $y_j = \sin(\xi_j)$, but now $x_j^b = \sin(\xi_j + \frac{\pi}{8})$, i.e. a sine wave with a phase shift. Plotted in Figure 9 are both the analysis vector and its FFT power spectrum for observations taken at (iv) every, (v) every other and (vi) every third grid point. In order to investigate the phase shift and exploit the methods explored in Section 2.1.5 the real and imaginary parts of the power spectrum for each example are plotted in Figures 10, 11, and 12 respectively. From these graphs several things can be seen.

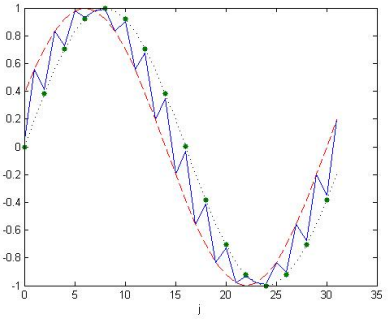
Firstly we consider example (iv). As can be seen in Figure 9(a) the analysis is a sine wave, but it has a small phase shift due to the influence of the background. The power spectrum for this analysis, shown in Figure 9(b), is as we would expect for a sine wave. It is worth noting the difference between this and example (i) which had a zero background $\boldsymbol{x^b} = 0$. Here there is no scaling factor on the analysis. By itself, however, Figure 9(b) does not tell us anything about the phase or the phase shift of



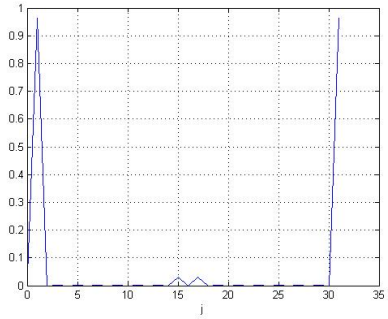
(a) The analysis x^a with observations taken every grid point.



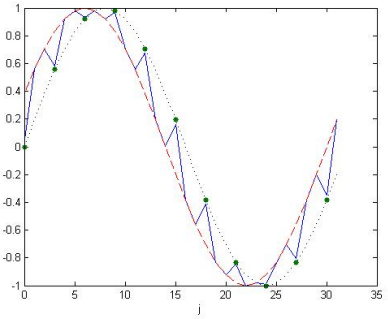
(b) The power spectrum where observations are taken every grid point.



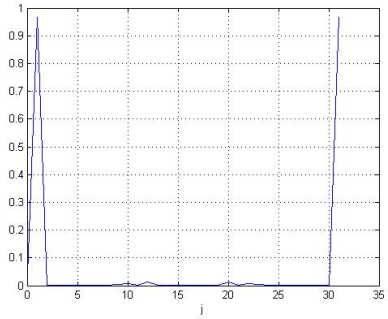
(c) The analysis x^a with observations taken every other grid point.



(d) The power spectrum where observations are taken every other grid point.

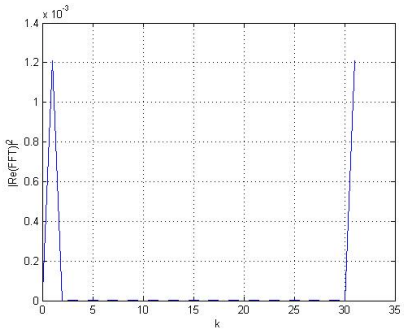


(e) The analysis x^a with observations taken at every third grid point.

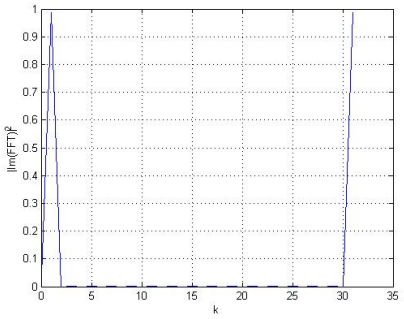


(f) The power spectrum where observations are taken every third grid point.

Figure 9: The analysis $\boldsymbol{x^a}$ are plotted as well as the power spectrum for each analysis. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and • show the observations.

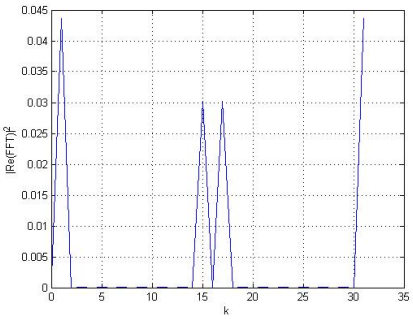


(a) The real part of the FFT

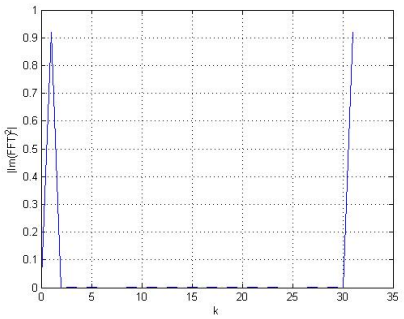


(b) The imaginary part of the FFT.

Figure 10: The real and imaginary parts of the FFT where observations are taken every grid point.

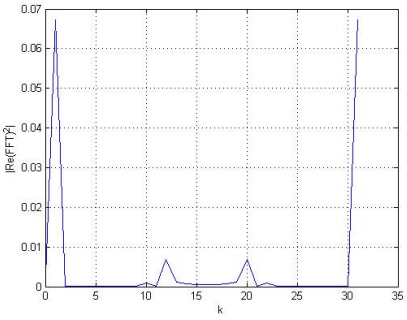


(a) The real part of the FFT

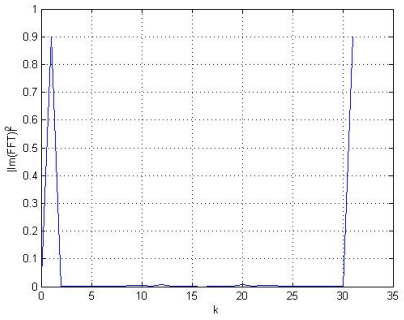


(b) The imaginary part of the FFT.

Figure 11: The real and imaginary parts of the FFT where observations are taken every other grid point.



(a) The real part of the FFT



(b) The imaginary part of the FFT.

Figure 12: The real and imaginary parts of the FFT where observations are taken every third grid point.

the analysis. To consider the phase we refer to Figure 10, which shows the real and imaginary parts of the FFT, and the methods explored in Section 2.1.5. The small phase shift in the analysis does show up in the real and imaginary power spectra. The amplitude in the imaginary part is slightly lower then that for a standard sine wave, but there is now a small amount of amplitude in the real part. Referring back to Section 2.1.5 this amplitude in the real part signifies a phase shift since a standard sine wave is a purely odd function.

Now we consider example (v), when observations are taken at every other grid point. We can see in Figure 9(c) that the analysis has the general shape of a sine wave but oscillates between the background and the observations. As in Section 3.2.1 this is due to the observations being weighted more heavily in the cost function, so at grid points where an observation is taken the analysis is very close to the observation value but at grid points where no observation is taken the analysis is set to the background, due to the matrix $\boldsymbol{B}$ being diagonal. There is a parallel here between examples (v) and (ii), shown in Figure 9(c) and Figure 6(c) respectively. In both examples the analysis oscillates about a sinesoidal shape, being set to the background $\boldsymbol{x^b}$ at any grid point where no observation is taken. There is also a parallel between the power spectra for these two examples. As there was in Figure 6(d), Figure 9(d) has two sets of peaks, one corresponding to the wave envelope and one corresponding to the oscillations.

This parallel between the analysis and the power spectra in the examples when $\boldsymbol{x^b} \neq \boldsymbol{0}$ and $\boldsymbol{x^b} = \boldsymbol{0}$ helps to confirm our reasoning on why they look the way they do, but what is perhaps more interesting is how and why they differ. When $\boldsymbol{x^b} = \boldsymbol{0}$ there is a scaling factor on the analysis for both the case of an observation at every and every other grid point, examples (i) and (ii) respectively. In contrast, when $\boldsymbol{x^b} \neq \boldsymbol{0}$ and in example (iv) we take observations at every grid point there is no reduction in amplitude of the power spectrum, but when in example (v) we take observations at every other grid point we can see in Figure 9(d) that the peaks at $k = 1, N-1$ are slightly lower than 1, which we know from Section 2.1.4 represents a scaling factor on our analysis. Also although for both examples (v) and (ii) the analysis is multiplied by a scaling factor, recognised by the reduction in amplitude at

$k = 1$ in Figures 9(d) and 6(d), there is a big difference in the size of this reduction. This is due to the wave envelope in each case. In Figure 6(c) we can see that the analysis is oscillating about a wave which has a scaling factor of about 0.45 where as in Figure 9(c) the oscillations are about a wave which is only scaled by a factor of about 0.95.

The peaks corresponding to the small scale oscillations also have a different amplitude when $\boldsymbol{x^b} \neq \mathbf{0}$ or $\boldsymbol{x^b} = \mathbf{0}$. This can again be understood by looking at Figures 9(c) and 6(c), which show the analysis for example (v) and example (ii) respectively. In Figure 9(c) the oscillations themselves are much smaller then they are in Figure 6(c). This difference in height of the oscillations is the reason for the difference in amplitude of the central peaks in the two power spectra shown Figures 9(d) and 6(d).

If we now look at the real and imaginary parts of the power spectrum, Figure 11 shows that as with example (iv), this analysis also has a phase shift. This can be seen by the lowered amplitude in the imaginary spectrum and the presence of amplitude in the real spectrum at $k = 1$. However, what is perhaps more worth noticing is that the peaks corresponding to the small scale oscillations are solely present in the real part of the spectrum. This distinction between the two waves present in the analysis and where their amplitudes lie in the real/imaginary spectra may prove useful in separating the wave we want from the unwanted small scale oscillations.

Lastly we consider the analysis for example (vi). As can be seen in Figure 9(e) the analysis again oscillates about a general sine wave shape. However, this time it is not a simple summation of distinct waves. This lack of clear wavenumbers present is the cause of the various small peaks in the centre of the power spectrum, Figure 9(f). It is also harder to separate the small scale oscillations from the wave we want because the amplitude in the FFT caused by the oscillations is from both the real and imaginary power spectra, as can be seen in Figure 12.

The small scale oscillations in the analysis when observations are taken at less than every grid point is caused by having to set the analysis to the background whenever there is no observation. This is because we are using an error covari-

Example Number	Value of $\boldsymbol{x^b}$	Observations at ... grid point
(i)	$\boldsymbol{x^b} = 0$	every
(ii)	$\boldsymbol{x^b} = 0$	every other
(iii)	$\boldsymbol{x^b} = 0$	only two
(iv)	$x_j^b = \sin(\xi_j + \frac{\pi}{8})$	every
(v)	$x_j^b = \sin(\xi_j + \frac{\pi}{8})$	every other
(vi)	$x_j^b = \sin(\xi_j + \frac{\pi}{8})$	every third

Table 1: A list of each example and its conditions.

ance matrix $\boldsymbol{B}$ which is diagonal, so observations can only affect the grid point at which they are taken. By using a matrix $\boldsymbol{B}$ with off diagonal entries we can spread the influence of a single observation and smooth the small scale oscillations in the analysis.

3.3 A Non-Diagonal Matrix $\boldsymbol{B}$

Instead of using a diagonal matrix $\boldsymbol{B}$ we now construct a matrix $\boldsymbol{B}$ with entries on the off-diagonals as well. It is important to note that $\boldsymbol{B}$ needs to be of a special form in order to preserve the periodic nature of our waveform. The sine wave is periodic, $\sin(\xi_j) = \sin(\xi_j + 2\pi)$, and although we have set up our domain so that only one of these end points is included ($j = 0, \ldots, N-1$), these periodic effects still need to be felt by the analysis and this is achieved by having a circulant[2] matrix $\boldsymbol{B}$. It is also worth noting here that $\boldsymbol{B}$ will be symmetric. The matrix $\boldsymbol{B} = \{B_{ij}\}$ now has the form

$$B_{ij} = \sigma_b^2 \exp\left(\frac{-\mu_{ij}\delta^2}{r^2}\right), \tag{18}$$

where

$$\mu_{ij} = \min\{|i-j|, N - |i-j|\},$$

[2]A circulant matrix is a matrix whose rows are composed of cyclically shifted versions of a list [3].

grid spacing $\delta = \frac{2\pi}{N}$ and r is the radius of influence[3]. Here the μ_{ij} term is a distance measurement which takes into account the periodic nature of our wave and the $\exp\left(\frac{-n\delta^2}{r^2}\right)$ term is a correlation function. The correlation function allows grid points close together to influence each other while leaving widely spaced grid points uncorrelated. It is based on but not identical to the correlation function described in [5]. Using symbols in order to make the pattern easier to see, our matrix $\boldsymbol{B}$ given by (18) can be written in the form

$$\boldsymbol{B} = \begin{pmatrix} \sigma_b^2 & \rho & \omega & \eta & & \dots & & \eta & \omega & \rho \\ \rho & \sigma_b^2 & \rho & \omega & \eta & \dots & & & \eta & \omega \\ \omega & \rho & \sigma_b^2 & \rho & \omega & \eta & \dots & & & \eta \\ \vdots & & & & & & & & & \\ \omega & \eta & & & & \dots & \eta & \omega & \rho & \sigma_b^2 & \rho \\ \rho & \omega & \eta & & & \dots & & \eta & \omega & \rho & \sigma_b^2 \end{pmatrix}. \tag{19}$$

With $\boldsymbol{B}$ no longer being simply a diagonal matrix, we can no longer use our simplified cost function. We must now use the cost function in its general form. From (9), with $h = \boldsymbol{H}$, we have

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x^b})^T \boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{x^b}) + \frac{1}{2}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x})^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}), \tag{20}$$

which is minimised by $\boldsymbol{x} = \boldsymbol{x^a}$, where

$$\boldsymbol{x^a} = \boldsymbol{x^b} + \boldsymbol{B}\boldsymbol{H}^T(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^T)^{-1}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x^b}). \tag{21}$$

The matrix $\boldsymbol{H}$ represents the linearised observational operator and is a p x n matrix whose entries are 0 or 1, where p is the number of observations and n is the length of the state vector $\boldsymbol{x}$.

We start by considering an example where we take a matrix $\boldsymbol{B}$ which has all its entries set to zero other than on the three central diagonals and the corner values. This is (19) with only σ_b^2 and ρ not equal to zero. We take observations at every other grid point and set $r = \delta$. As before $y_j = x_j^t = \sin(\xi_j)$, $x_j^b = \sin(\sigma_j + \pi/8)$ and $\sigma_b^2 = 1$, $\sigma_o^2 = 0.1$. The analysis generated from this example is shown in Figure 13. As can be seen the analysis still oscillates but comparison with Figure 9(c), which

[3] r is a length scale and should have the same dimensions as δ. Therefore $r =$'1 gridspace' corresponds to $r = \delta$ and $r =$'α gridspaces' corresponds to $r = \alpha\delta$, where $\alpha = 1, 2, 3, \dots$

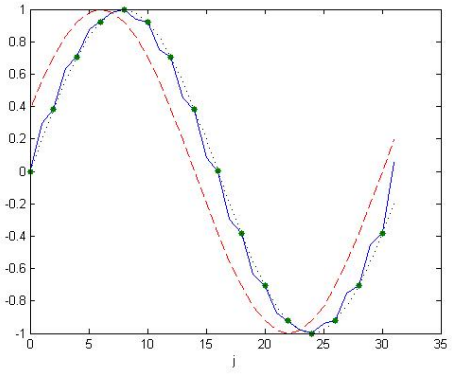
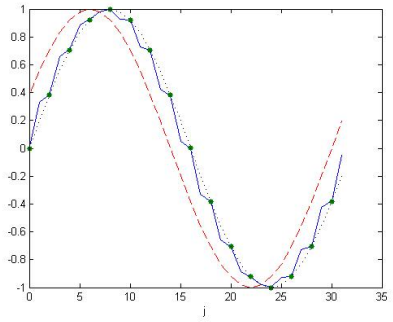


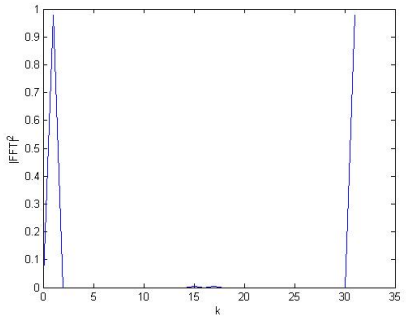
Figure 13: The analysis $\boldsymbol{x^a}$ using a tridiagonal matrix $\boldsymbol{B}$ with radius of influence $r = \delta$ and observations taken at every other grid point. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and $\bullet$ show the observations.

shows the analysis when a diagonal matrix $\boldsymbol{B}$ is used, shows that this oscillation has been significantly reduced. This is due to the matrix $\boldsymbol{B}$ spreading out the observational information and allowing grid points where no observation is taken to be affected by the observations at neighbouring grid points.

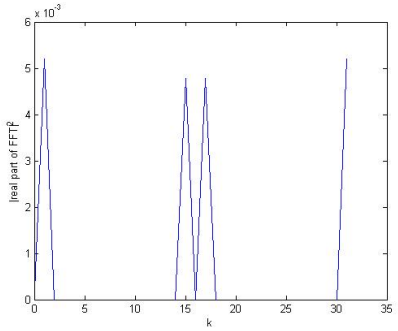
Keeping everything as in the case above, we now consider the analysis generated when a full matrix $\boldsymbol{B}$, as defined in (18), is used. Figure 14(a) shows the analysis when the matrix $\boldsymbol{B}$ has radius of influence $r = \delta$. With $r = \delta$ an observation at a grid point can affect the analysis at its neighbouring grid points. This is the same range of influence as we had with the tridiagonal matrix $\boldsymbol{B}$ and comparison between Figures 14(a) and 13 show that we do get qualitatively similar analyses in both cases. Figure 14(b) shows the power spectrum of the analysis. As can be seen this power spectrum is very similar to that which we would expect for a sine wave with large peaks at $k = 1, 31$ of amplitude almost equal to one. There are also small peaks at $k = 15, 17$ corresponding to the wavelength of the small scale oscillations. By considering the real and imaginary parts of the FFT, shown in Figures 14(c) and 14(d), and comparing them with those for when a diagonal matrix $\boldsymbol{B}$ is used, Figures 11(a) and 11(b), we can see that the phase of the analysis is also better corrected by a matrix $\boldsymbol{B}$ which has entries everywhere compared to one which is diagonal. The amplitude in the real part at $k = 1$ is smaller for the full matrix $\boldsymbol{B}$



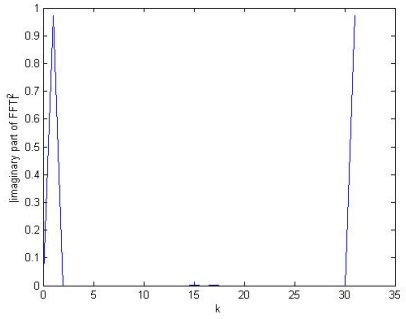
(a) The analysis $\boldsymbol{x^a}$



(b) The power spectrum.



(c) The real part of the FFT



(d) The imaginary part of the FFT.

Figure 14: The analysis $\boldsymbol{x^a}$ using a full matrix $\boldsymbol{B}$ with radius of influence $r = \delta$ is plotted as well as the power spectrum and the real/imaginary part of the power spectrum. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and • show the observations.

case showing that the phase shift is smaller and that the analysis is closer to the truth.

The full matrix $\boldsymbol{B}$ with radius of influence $r = \delta$ generates qualitatively the same analysis as the tridiagonal matrix $\boldsymbol{B}$. In order to achieve a smoother analysis we need to increase the radius of influence in the matrix $\boldsymbol{B}$. When $\boldsymbol{B}$ has a larger radius of influence the observational information is spread out further and can affect more grid points. This spreading of observational information generates a smoother analysis. Figure 15 shows the analysis when the matrix $\boldsymbol{B}$ has a radius of influence $r = 3\delta$. As can be seen this analysis is much smoother than the one in Figure 14(a). This is because each observation is now affecting several grid points and therefore the information is more evenly spread, creating a smoother analysis.

So far all of the calculations have been done using perfect observations. However,

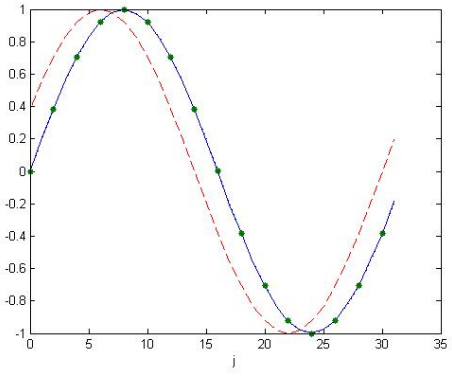


Figure 15: The analysis $\boldsymbol{x^a}$ using a full matrix $\boldsymbol{B}$ with radius of influence $r = 3\delta$. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and • show the observations.

in reality all the observations we have will have some form of error associated with them. We now go on to consider the analysis generated when the observations have errors.

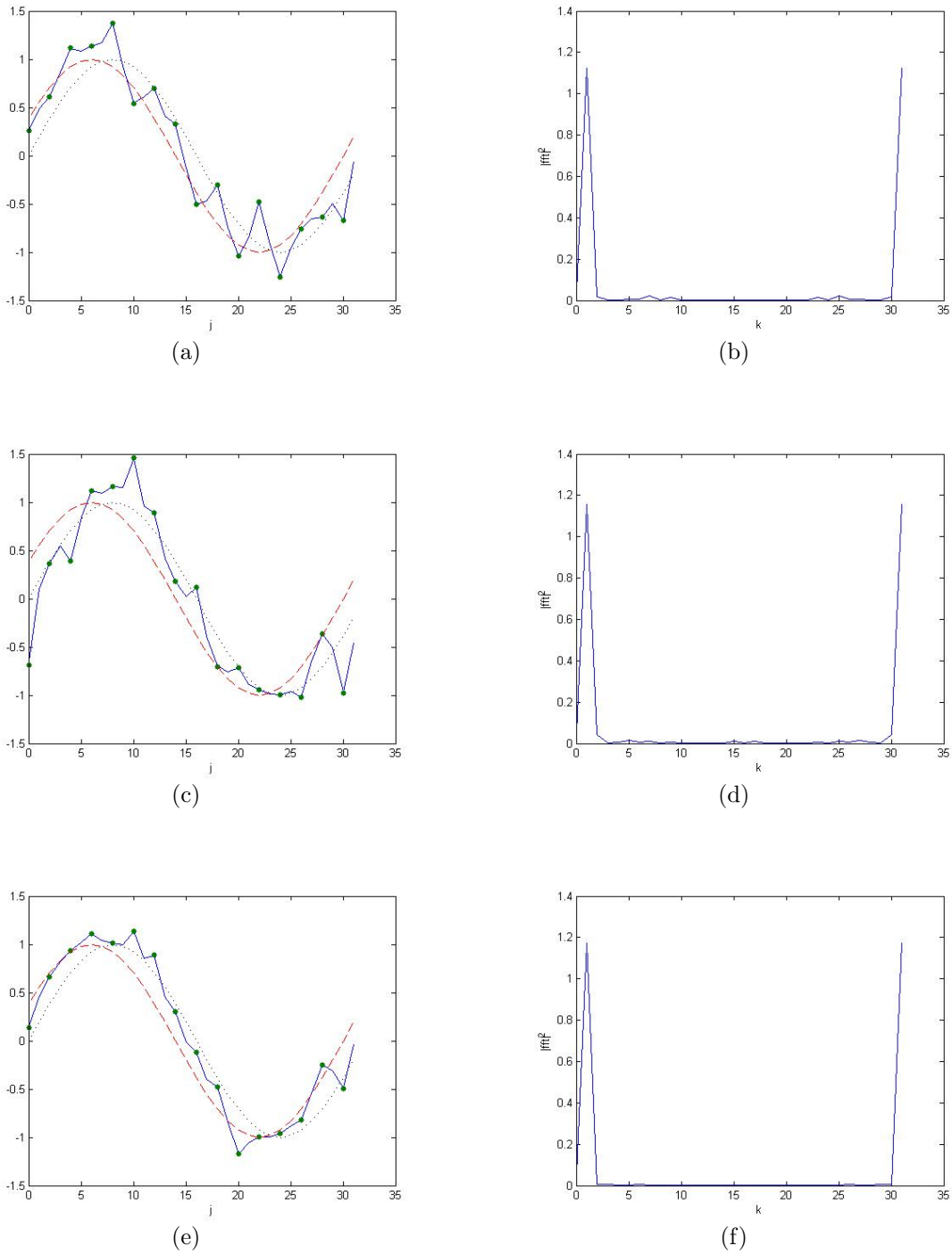
4 Observations with Errors

We now consider observations which have error. We generate our observations by taking

$$y_j = x_j^t + \text{random noise}$$

for $j = 0, \ldots, N-1$. In other words we take perfect observations from the truth and then add random noise. This random noise has a normal distribution with mean zero and variance σ^2. Everything else is left as it is for the case in Section 3.3 where a full matrix $\boldsymbol{B}$ with radius of influence $r = \delta$ is used; $\sigma_b^2 = 1$, $x_j^t = \sin(\xi_j)$, $x_j^b = \sin(\xi_j + \pi/8)$ and observations are taken at every other grid point. However, instead of choosing σ_o^2 to be arbitrary it is now taken to be equal to the variance of the error on the observations, $\sigma_o^2 = \sigma^2$.

In order to investigate the effect the random error has on the analysis we generate three different sets of random error each with the same variance. We then calculate the analysis in each case. The analyses for a random error with variance $\sigma^2 = 0.1$ are plotted in Figure 16. The power spectrum of each analysis is also shown. By comparing these plots with the analysis and power spectrum generated with perfect observations, Figures 14(a) and 14(b), we can see that the random error has a big effect on both the analysis and the power spectrum. However these effects manifest differently for the two types of plot. Considering Figures 16(a), 16(c) and 16(e), we can see that the different random error causes the three analyses to have considerable differences between them. The locations of the small scale oscillations differ as well as their amplitudes. Despite these differences all three do have a common feature, they have all lost the regular pattern that characterised the analysis with perfect observations in Figure 14(a). Although the analyses differ, the power spectra for the three cases are remarkably similar. Figures 16(b), 16(d) and 16(f) all have small amplitude fluctuations at various wavenumbers corresponding to the noise in the individual analysis, but all three have the large peaks at $k = 1$ associated with a sine wave. This shows that despite the error on the observations the cost function still generates a sinusoidal analysis. The amplitude of the peak at $k = 1$ is also very similar in all three cases. This peak is greater than one which, from what we saw in Section 2.1.4, shows the error is causing a scaling factor to act on the analyses.



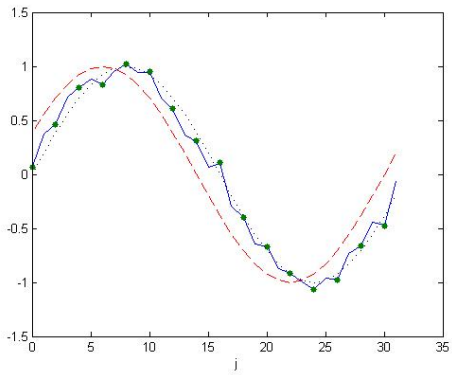
(a) (b) (c) (d) (e) (f)

Figure 16: The analysis $\boldsymbol{x^a}$ for three different sets of random error all with variance $\sigma^2 = 0.1$ are plotted as well as the power spectrum for each analysis. A full matrix $\boldsymbol{B}$ is used with radius of influence $r = \delta$. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and • show the observations.

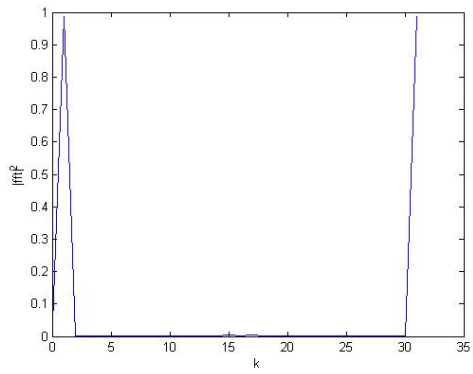
This can be seen by looking at Figure 16(a). The cost function is weighted to give most influence to the observations so any error on the observations will greatly effect the analysis. Here the error on the observations is increasing the amplitude of the analysis envelope which causes an increase in the amplitude of the peak in the power spectrum.

We now consider the analyses for a random error with a smaller variance, $\sigma^2 = 0.01$. Now that $\sigma^2 = 0.01$, we also set $\sigma_o^2 = 0.01$ but σ_b^2 is left unchanged. The analyses for three different sets of random numbers all with variance $\sigma^2 = 0.01$ are plotted in Figure 17. The power spectrum of each analysis is also shown. As was the case where $\sigma^2 = 0.1$, the three analyses shown in Figures 17(a), 17(c) and 17(e) vary from each other depending on their own random error. The magnitude of the oscillations has been reduced due to the reduction in the size of the error. If we consider the power spectra shown in Figures 17(b), 17(d) and 17(f) we can see that all three are very similar. As in Figure 16 where $\sigma^2 = 0.1$, all three have small amplitude fluctuations at various wavenumbers corresponding to the noise of the individual analysis but again all three have the large peaks at $k = 1$ associated with a sine wave. The amplitude of the peak at $k = 1$ is however now much closer to one which is what we want for a normal sine wave. This reduction in amplitude of the power spectrum peaks in Figure 17 compared to Figure 16 can be understood by looking the plots of the individual analyses. As we can see the amplitude of the wave envelope in Figure 17 is much closer to what we expect for a normal sine wave. This is due to the error on the observations having a lower variance. It is this smaller error variance that causes a more acurate analysis and a power spectrum peak amplitude closer to one.

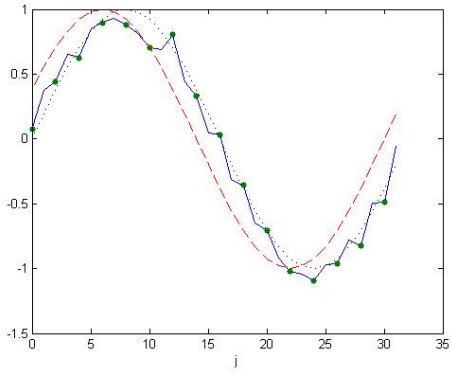
In order to examine the power spectrum more closely and consider the phase of the analysis the real and imaginary parts of the power spectrum are plotted in Figure 18. Due to all the three power spectra in Figure 17 being so similar, we choose to do this just for the analysis shown in Figure 17(e). As we can see in Figure 18(a) the real part of the FFT is extremely small, $O(10^{-3})$. This indicates that despite the error on the observations the cost function is still correcting the phase shift. By also looking at Figure 18(b) we can see that all of the amplitude corresponding to



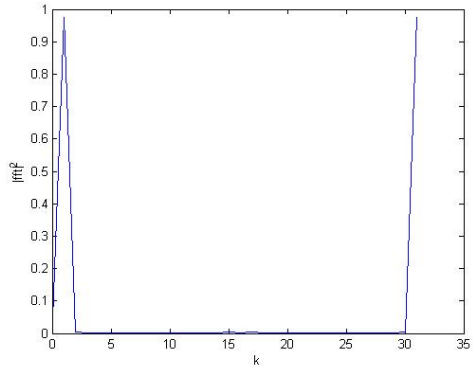
(a)



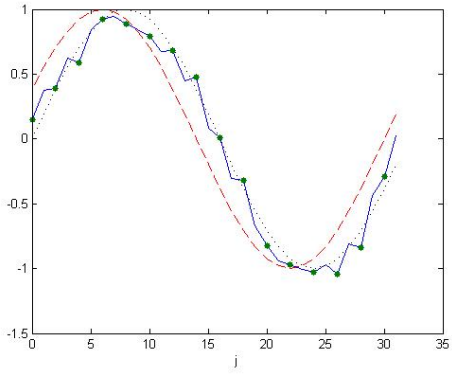
(b)



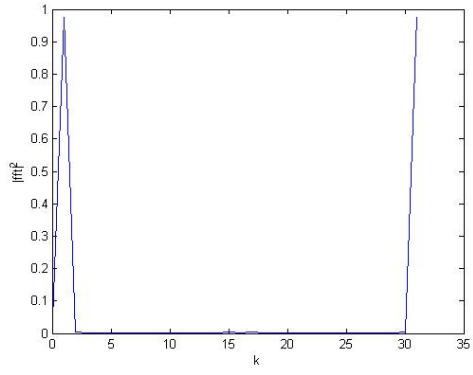
(c)



(d)

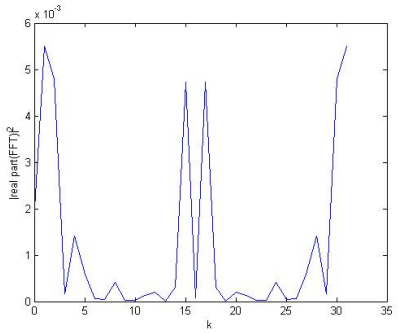


(e)

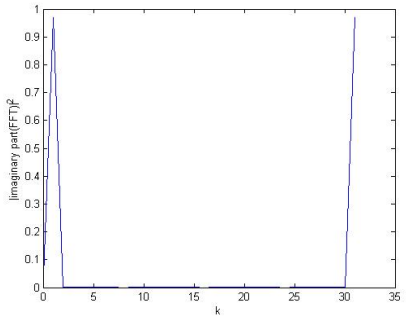


(f)

Figure 17: The analysis $\boldsymbol{x^a}$ for three different sets of random error all with variance $\sigma^2 = 0.01$ are plotted as well as the power spectrum for each analysis. A full matrix $\boldsymbol{B}$ is used with radius of influence $r = \delta$. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and $\bullet$ show the observations.



(a) The real part of the FFT



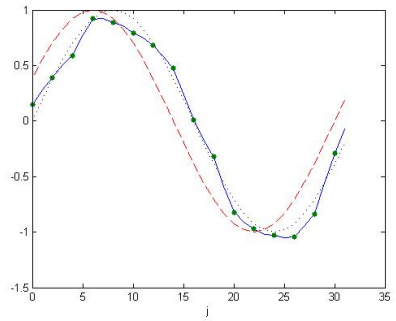
(b) The imaginary part of the FFT.

Figure 18: The real and imaginary parts of the power spectrum shown in Figure 17(f). The observations have random error with variance $\sigma^2 = 0.01$ and a full matrix $\boldsymbol{B}$ is used with radius of influence $r = \delta$.

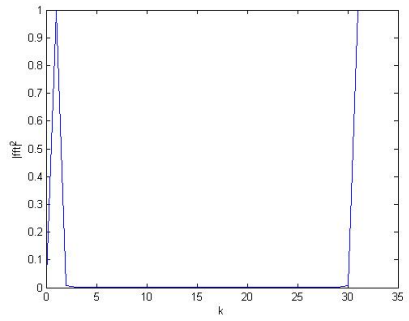
the small scale oscillations is located in the real part of the FFT. This agrees with what we saw using perfect observations and may be a way to separate the noise from the wave we want. It is also worth noting that although the small scale oscillations in the analysis are no longer regular their wavenumber still shows up in the real part of the power spectrum, evidenced by the higher peaks at $k = 15, 17$ in Figure 18(a).

To investigate further the effects of observation error we now repeat the experiments using the same random error as was used in case 17(e) where $\sigma^2 = 0.01$, but now we use a full matrix $\boldsymbol{B}$ with a radius of influence $r = 3\delta$. The analysis, power spectrum and real/imaginary power spectra are shown in Figure 19. As we can see the analysis is much smoother than it was in Figure 17(e). This is because the matrix $\boldsymbol{B}$ now has a radius of influence $r = 3$ so spreads out the observation information further, resulting in a smoother analysis. The power spectra are very similar to the ones shown in Figures 17(f), 18(a) and 18(b). In Figure 19(b) the amplitude at $k = 1$ is even closer to one but the main difference can be seen in the real part of the FFT shown in Figure 19(c) which does not have peaks at $k = 15, 17$, as were seen in Figure 18(a). This is because the analysis has been smoothed by $r = 3$ in the matrix $\boldsymbol{B}$ and so no longer exhibits the recognisable small scale oscillations.

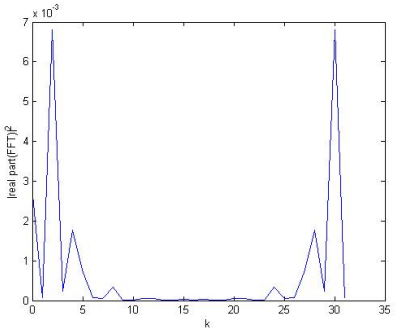
So far we have always assumed that the observations are more accurate than the background and have consequently weighted the observations more heavily in the cost function so as to give them more influence over the analysis. Now we consider the case where the uncertainty in the observations is equal to that in the



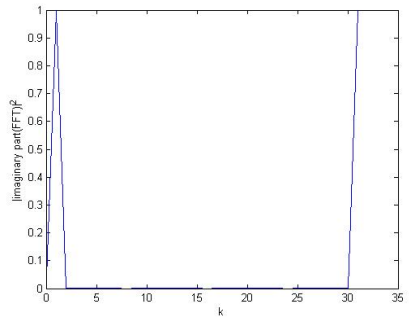
(a) The analysis $\boldsymbol{x^a}$



(b) The power spectrum.

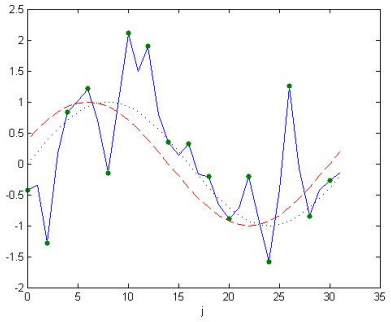


(c) The real part of the FFT

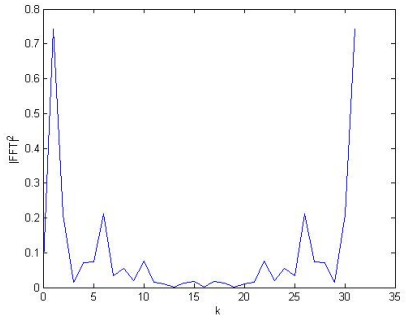


(d) The imaginary part of the FFT.

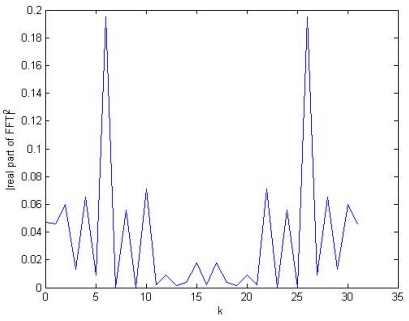
Figure 19: The analysis $\boldsymbol{x^a}$ is plotted as well as the power spectrum and the real/imaginary part of the power spectrum. The observations have random error with variance $\sigma^2 = 0.01$ and a full matrix $\boldsymbol{B}$ is used with radius of influence $r = 3\delta$. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and $\bullet$ show the observations.



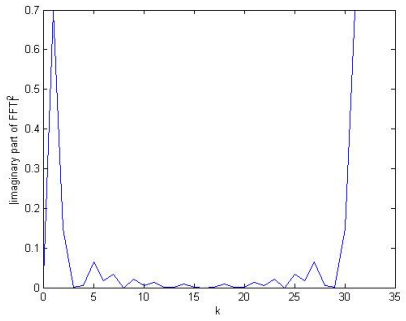
(a) The analysis $\boldsymbol{x^a}$



(b) The power spectrum.



(c) The real part of the FFT



(d) The imaginary part of the FFT.

Figure 20: The analysis $\boldsymbol{x^a}$ is plotted as well as the power spectrum and the real/imaginary part of the power spectrum. The observations have random error with variance $\sigma^2 = 1$ and a full matrix $\boldsymbol{B}$ is used with radius of influence $r = \delta$. The solid line is the analysis, the dashed line is the background, the dotted line is the truth and $\bullet$ show the observations.

background. We set the variance of the random error on the observations to be $\sigma^2 = 1$ so then $\sigma_o^2 = 1 = \sigma_b^2$. We still take observations every other grid point and a radius of influence $r = \delta$. The analysis and its power spectra are shown in Figure 20. This example shows how useful the FFT is. By simply looking at the analysis shown in Figure 20(a) we see that it oscillates wildly with very little resemblance to the sine wave at all. We could therefore conclude that the error on the observations has overpowered the true signal and that the analysis has no skill. However when we look at the power spectrum shown in Figure 20(b) we see that actually the analysis does have some skill. The large peaks at $k = 1, 31$ that we expect for a sine wave are clearly visible. Their amplitude is lower than we would get for a normal sine wave so our analysis has a scaling factor but it is still dominated

by a sine wave. If we now also consider the real and imaginary parts of the FFT, shown in Figures 20(c) and 20(d), we also see that the majority of the amplitude at $k = 1$ is in the imaginary part of the spectra, which is what we want for a sine wave, so the phase shift in the background has also been partially corrected. These results are encouraging as they suggest that we can get some skill in the analysis even from observations with a large error. They also show that it can be misleading to judge the analysis by simply looking at its numerical values.

From the results shown in Section 4 we have seen that we can learn about an analysis generated using imperfect observations using the methods developed in Section 2. We are still able to investigate amplitude, wavenumbers and phase just as we can when the observations are perfect. We are also able to achieve results when our observations have small errors that are very similar to those we achieve with perfect observations. This is good news as in reality all our observations will have some form of error associated with them. In reality we never have perfect observations.

5 Conclusions and Future Work

We have shown how the DFT and its power spectrum can be used to identify properties of a wave. The wavenumber, scaling factor and phase can all be identified by applying the DFT. This is particularly useful for more complex waveforms where the properties of the wave are not clear from the waveform itself.

We then went on to apply the DFT techniques to some simple variational analyses. The DFT allowed us to examine the analysis and to understand better the effects of the observational information. The power spectrum allowed us to separate the small scale oscillations in the analysis from the smooth wave we wanted. The power spectrum also enabled us to identify and differentiate between phase and amplitude errors. By considering the real and imaginary parts of the DFT separately we are able to examine the phase of the analysis and also the phase shift in comparison to the truth. This allows us to consider not only whether the analysis reconstructs the correct wavenumber and wave shape but also whether the analysis has the correct phase.

We generated analyses using both perfect observations and observations with error. These provided encouraging results. The analyses generated when the observations had error did not look promising as they oscillated wildly and did not look like a sine wave but by using the DFT we saw that they actually were a reasonable reconstruction of the truth. This is encouraging because in reality we will never have perfect observations.

Throughout this work we kept $\sigma_b^2 = 1$ and varied σ_o^2. In the future it would be interesting to try varying σ_b^2 as well as σ_o^2 to see if there is an optimal balance that can be reached. We could also use a background that is a sine wave with random noise so that we know the variance σ_b^2 as well as σ_o^2.

Once we have finished with this one wavelength domain the next step is to consider a wave which is longer than our 1D domain. We want to investigate whether the same techniques that have been used here can be applied when only part of the wave is contained in the domain. The matrix $\boldsymbol{B}$ will need to be considered in particular as the wave will no longer be periodic within the model domain.

Acknowledgements

This work was funded by the UK Natural Environment Research Council (NERC) through a Data Assimilation Research Center CASE studentship in collaboration with the Met Office. The authors are grateful to S.P. Ballard from the Met Office for useful discussions and comments.

References

[1] Arfken, G. (1985) Mathematical Methods for Physicists. Academic Press Inc.

[2] Cizek, V. (1986) Discrete Fourier Transforms and their Applications. Adam Hilger.

[3] Davies, P.J. (1979) Circulant Matrices. Wiley.

[4] Hesthaven, J.S. and Gottlieb, D. (1998) Spectral Approximation of Partial Differential Equations: Numerical Analysis and Applications. Lecture Notes, Division of Applied Mathematics, Brown University, Providence, RI, USA.

[5] Kalnay, E. (2003) Atmospheric Modeling, Data Assimilation and Predictability. Cambridge University Press.

[6] Lawless, A.S. and Bannister, R. (2006) Theory and Techniques of Data Assimilation MAMB10. Lecture Notes, University of Reading.

[7] Nichols, N.K. (2003) Data Assimilation: Aims and Basic consecpts. Data Assimilation for the Earth System, (eds. R. Swinbank, V. Shutyaev, W.A. Lahoz), Kluwer Academic, pp. 9-20.

[8] Ramirez, R. W. (1985) The FFT Fundamentals and Concepts. Prentice-Hall.